IMAGERY, COGNITIVE PENETRABILITY, AND THE LOCATION OF CONTENT.

A Thesis
submitted in partial fulfilment
of the requirements for the degree
of
Master of Arts in Psychology
in the
University of Canterbury
by
Charles Heywood.

University of Canterbury

1990

CONTENTS

ABSTRACT.

Recent results from the theory of meaning and mental content are applied to Zenon Pylyshn's criticism of the 'Analog' theory of quasi-spatial image representations on the grounds of the 'cognitive penetrability' of the purported image manipulation tasks. The latter condition asserts that if behaviour alters in such a way as to correlate with the 'meaning' or reference of sensory inputs, as is the case in imagery tasks, then the explanation of the underlying cognitive function must appeal to operations upon sentence-like 'symbolic ' representations, as opposed to the 'quasi-spatial medium' proposed by the Analog school. This follows from the assumption that relations between contentful semantically interpreted constructs like belief must be mirrored by, and thus explained in terms of, correlative changes at the physically instantiated, alogarithmic, 'symbol level'. Because the explanatory construct utilised at this level is symbolic content, cognitive processes identified as 'symbolic' by dint of their alteration in concert with environmental stimuli, or the objects of cognition, are thereby, allegedly, only to be explained in terms of symbolic processes, which are instantiated by but not equivalent to physiological processes. Counter-instances from the philosophical literature are recounted wherein the contents of physically constant mental states are seen to alter with environmental circumstance, in violation of the cognitive penetrability condition. With the latter invalid the weight of empirical evidence supports the Analog position. Stich's 'Syntactic Theory of Mind' is suggested as a general theoretical and philosophical framework for the Analog Theory.

INTRODUCTION.

FROM RATIONALISM AND NATURALISM TO INTERNALISM AND EXTERNALISM.

What makes Cognitive Psychology cognitive is the centrality of intentional notions like representation, content, and information (in its non-technical sense) to the explanation of rational behaviour. For some time during the sixties and seventies it seemed as if Psychology had come full circle since the lean Behaviourist years to implicitly reaffirm Brentano's adage that 'intentionality is the mark of the mental'. But even as Fodor described "your standard contemporary cognitive psychologist" as wedded to the idea that "mental states are distinguished by the *content* of the associated representations" in 1980, he was calling for the abandonment of this kind of psychology in favour of a purely formal "Methodological Solipsism" which specifically disavows intentional idioms.[1]

The problem is that whilst beliefs, desires, hopes, memories, attitudes and the like seem intimately tied up with what it is to be human and to act intelligently on the basis of evidence, none of those internal states of our heads which have been posited in order to explain cognitive behaviour seem to be the sorts of things of which intentional states can be predicated. My belief that 'snow is white' for instance, has truth conditions, - it may be true or false, - but those internal states whose causal efficacy is dependent upon their form, like computation, can be no more true or false than a pebble in my garden. As Fodor puts it:

> computational processes are both *symbolic* and *formal.*

Yet,

---

[1] Fodor (1980), p.64.

> Formal processes are the ones that are specified without reference to such semantic properties of representations as, for example, truth, reference and meaning.[2]

However, if all cognitive processes are in the last analysis brain processes it would seem that *all* intra-personal theories of cognition must be essentially formal (= achieving their effect mechanically, by virtue of their form or structure).

The result is a paradox; what we *ordinarily* think of as mental states seems incompatible with the best theories available to explain them. Mental states, so conceived, possess the quality of *intentionality* (meaning, content, semantics, representation or reference) which cannot be possessed (so the argument goes) by the scientific theoretical entities of Cognitive Psychology. Consequently these intentional states, like my belief that 'the world is round', cannot be 'in the head' where the theoretical entities presumably reside.

Three resolutions seem to be available; either we bite the bullet and be prepared for the probability that 'Cognitive Psychology' as presently understood will be revised into a purely formal 'Methodologically Solipsist' discipline which concerns itself solely with what is demonstrably 'in the head' and has little or nothing to do with what we ordinarily conceive of as thought, belief or desire, or indeed with the world outside the head (cf Stich (1978), (1983), Fodor (1980)); or we defend to the last Brentano's ideal by asserting that if formal-computational psychology is inconsistent with propositional attitude psychology (psychology which attributes beliefs and desires), then so much the worse for the former (cf Churchland (1979), Burge (1982), (1986)); or we can take the third path and assert either that the contradiction between formality and intentionality may be reconciled and a compromise reached (e.g. McGuinn (1989)), or the problem may be discounted with little discussion (Pylyshn (1984)).

This issue has come to preoccupy that domain which is simultaneously the philosophical side of cognitive psychology and the practical, scientifically directed, side of the philosophy of mind. This comes as no surprise for representation, and thus mental content, are pivotal in modern Psychology. The question being asked above is the question as to whether content can be determinate in the head over and above dispositions to behave or relations to external objects, as opposed to the

---

[2] *ibid* .

thesis that psychological events can only be individuated in a way that makes necessary reference to external objects and relations, such that they constitute, in part, the psychological events in question, which is simultaneously the question as to *where mental content is located , what mental content is* , and thus *what psychological events are.* Yet as is so often the case in Science, the practising cognitive psychologist continues to use the concept of representation as his immediate situation demands, with little concern for the fate that may have been decided for him in journals he doesn't read.

The idea that the mind is determinate and solely 'within the head' can be traced back to the Rationalist philosophy of the seventeenth century.

Descartes founded this thesis with arguments that retain a great deal of persuasiveness today : Is it not possible that I may be deceived somehow about the deliverances of my senses ? Thus, all I can be certain of is what I am immediately aware of. It may be that nothing else exists apart from what I am immediately aware of - my Ideas - consequently my Ideas must be able to exist without the rest of the world, consequently they have an identity or nature independent of the external world. From an epistemological point:-

> ... The point is that in considering these arguments we come to realise that they are not as solid or as transparent as the arguments which lead us to knowledge of our own minds and of God, so that the latter are the most certain and evident of all possible objects of knowledge for the human intellect.[3]

- comes an ontological point:

> the mere fact that I can clearly and dictinctly understand one substance apart from another in enough to make me certain that one excludes the other.[4]

Thus the fact that the mind is better known than things beyond it leads into its being a different and distinct substance. Reed (1982) and Fodor (1980) are amongst those who have cited the tacit influence of Descartes'

---

[3] Descartes *Meditations* , p.75 Cottingham et al.
[4] Descartes *Objections and Replies* , p.146 Cottingham et al.

ideas on modern psychology. Even as the latter congratulates itself as being a 'new' science untainted by old philosophical dogmas, there remains a tendency, particularly in the psychology of perception, to view "all awarenessess (as) awarenesses of brain states ... as opposed to, say, awarenesses of objects ...".[5] A powerful rationale thus presents itself: if immanent cognitive activity has its own integrated and intrinsic form by virtue of being immediate and physical, the mind may be studied in isolation from the environment, and its intentional values recovered, by and large, from its own intrinsic principles. As Neisser has put it in connection with perception, raw perceptual input is rendered meaningful by 'processing, processing and more processing'. Burge terms this sentiment 'Individualism':

> According to individualism about the mind, the mental natures of a person's or animals mental states (and events) are such that there is no necessary or deep individuative relation between the individual's being in states of these kinds and the nature of the individual's physical or social environments.[6]

In a similar vein, McGinn characterises 'Internalism' as the idea that mental states are in the head where they are "essentially independent of the surrounding world".[7] And such, according to Fodor, are the views of "all cognitive psychologists except Gibsonians".[8] If one happens also to believe that Psychology is properly the science of consciousness - a dominant view until this century - rather than representation, the risk seems to be courted of removing Psychology from the physical world altogether. As James remarked in 1904:

> At first, 'spirit and matter', 'soul and body', stood for a pair of equipollent substances quite on a par in weight and interest. But one day Kant undermined the soul and brought in the transcendental ego, and ever since then the bipolar relation has been very much off its balance.[9]

---

[5] Reed (1982), p.733.
[6] Burge (1986) in Silvers (1989) p.39.
[7] McGinn (1989), p.1
[8] Fodor (1984) in Silvers (1989) p.1
[9] James (1904), in McDermott (1967), p169.

James, along with Pierce and Dewey, deplored the Dualism of Rationalist philosophy and Introspectionist Psychology that set the core of human life so much apart from its worldly content. In the book which more or less founded modern Psychology James wrote:

> On the whole, few recent formulas have done more service of a rough sort in Psychology than the Spencerian one that the essence of mental life and of bodily life are one, namely, the adjustment of inner and outer relations. Such a formula ... takes into account the fact that minds inhabit environments which act on them and on which they in turn react ...[10]

This is echoed by the predominant theme in American Naturalist Philosophy which emphasises the holism of man-in-the-world and the continuity of Science and Philosophy over the cerebral concerns of the Rationalism it replaced. Here Dewey reproaches the Rationalists for turning the virtue of subjective clarity into a vice:

> Philosophical dualism is but a formulated recognition of an impasse in life ... refugee idealism based on rendering thought omnipotent in the degree in which it is ineffective in concrete affairs:- these forms of subjectivism register an acceptance of whatever obstacles at the time prevent the active participation of the self in the on-going course of events. Only when obstacles are treated as challenges to remaking of personal desire and thought, so that the latter integrate with the movement of nature and by participation direct its consequences, are opposition and duality rightly understood.[11]

This picture is not difficult to understand: an organised taxonomy of consciousness and subjective mental content, if approached from Descarte's epistemological standpoint, can only really be achieved in an armchair with one's eyes closed, with all that this connotes for objectivity.

---

[10] James (1890), quoted in Fodor (1980), p.64.
[11] Dewey (1925), p.241.

However, the Psychology of the time which accompanied these suggestions - Behaviourism - went far beyond informed compromise and simply rejected mental constructs and Philosophical awareness to boot.[12] Rather than allowing internalistic themes to participate in the gradual emergence of what at the time was a very young science, the 'Mind' was jettisoned as excess ballast. The semi-Philosophical area of 'High-Theory'which might be employed in the study of content failed to determine subject matter, rather, it was determined - as irrelevant.

The Behaviourist vision is altogether different from modern 'Externalism'. For a start the latter can only be said to have existed in its present form since the late seventies. Secondly; unlike Behaviourism it does not forgo representation, but is itself an approach to representation. The sense of this is aptly given by a quote of Gibson's:

> there is no content of awareness independent of that (thing) of which one is aware.[13]

Externalism is thus an individuative thesis for mental states which holds the individual's environment to be constitutative of the nature of his mental states. It is thus a broadly naturalistic approach to issues of content which eschews the extremes of Behaviourism whilst opposing Cartesian currents in modern psychological theory. Thirdly; it is sufficiently esoteric to be little voiced in the mainstream Psychological Literature. The evidence upon which this view depends will be presented at the appropriate point below.

My interest is in the practical application of certain externalist philosophical themes to theoretical debate in Psychology, which their application may illuminate and decide. This brings us to the 'Imagery Debate'.

This is, or was, a debate about mental representation in a given domain. Interest was aroused by a flurry of papers in the 1970's which seemed to support an almost naive or Lockean conception of images as 'pictures in the head', which garnered a predictable reply. However by the mid-eighties most of the gambits had been played and the possibilities for new experimentation waned whilst the original protagonists simply retrenched their positions. Of course each side had 'conclusive' reasons for

---

[12] see Watson (1914).
[13] Gibson (1979), p.239.

assuming victory, but these were by-and-large partisan, and did not incorporate anything resembling a new result from outside the arena.

I will be taking another look at these issues , for the question as to whether an 'image' is in fact something picture-like or something sentence-like is possibly the single most marked disagreement on questions of representation and content that can be conceived. Large differences in the competing forms of representation provide a prima facie leverage for a decision procedure given the independent merits of Internalism and Externalism. If images are literally pictures of-themselves, then their representational characteristics and hence content are relatively unmysterious aspects of physical structure. If something like sentences or propositional representations are involved, however, like sentences of written English, they will possess something like 'meanings' which connect with their external objects *by virtue of* these meanings. 'Dog' doesn't look like a Dog, it just means Dog. It has something of 'Dogness' in it. It seems to possess a content-nature, and in no readily apparent physical way, which is of course internal 'to the head' on the grounds that the sentence itself is in the head A picture of a Dog, in contrast - scratched in the sand say - possesses its content by virtue of a relation to an object external to the picture itself. It has content in a physically palpable way; we don't feel the need to ascribe 'content-nature' to a section of beach.

As I am primarily concerned with the theoretical overview I will concentrate my descriptions on a suitable exemplar for each school. The issues are in any case well known and well publicised. I will also forgo a historical introduction of disagreements about imagery over the ages for two reasons: Firstly the imagery issue is a vehicle, a case in point rather than an end in itself. Similar arguments to those I will deploy could perhaps induce theoretical change in various avenues of cognitive psychology. Secondly: The contemporary structure of the imagery issue is quite distinct. The image-phile has never before been able to allude to empirical evidence. The sentiments which, say, Ryle voiced in the 1940's are only incidentally similar to those voiced by Pylyshn in the 1980's. The arguments are discontinuous. Nothing much happened of interest to us until 1971.

I. THE CONTEMPORARY LITERATURE.

There was little more than subjective evidence for the pictorial quality of mental images until Shepard and Metzler published their now classic report "Mental Rotation of Three Dimensional Objects" in 1971. Their findings, and reaction to their findings, established a precedent for academic discussion of this topic. In a series of experimental papers those influenced by Shepard and Metzler would continue to apply methodological ingenuity to the establishment and clarification of an autonomous domain of picture-like or "Analog" processes throughout the seventies and early eighties, whilst those influenced by certain foundational assumptions about cognitive representational processes would continue to author in the main purely theoretical rebuttals of the alleged significance of these experiments.[14]

The central aim of this thesis will be to provide a theoretical response to these theoretical arguments. As my case does not depend upon a close scrutiny of the experimental evidence the review presented below is expository rather than comprehensive or critical.

Shepard and Metzler had observed that:

> Human subjects are often able to determine that two-dimensional pictures portray objects of the same three-dimensional shape even though the objects are depicted in very different orientations.[15]

With this in mind they set out to measure the time course of this processing function. Subjects were given some 1600 exposures to line drawings of pairs of tetrahedrons at different orientations with respect to each other . Each subject was asked to respond as quickly as possible as soon as he or she was able to ascertain that the two line drawings were different depictions of the same object. (A certain percentage of the pairs were un-matching distractors). Surprisingly, over a total of 12800 trials a remarkably consistent linear relationship pertained between the angular

---

[14] For comprehensive reviews of the Analog literature see Kosslyn, Pinker, Smith and Shwartz (1979), Kosslyn (1980) and Shepard and Cooper (1982). For a full version of the contrary argument see Pylyshn (1984).

[15] Shepard and Metzler (1971), p. 701.

disparity of each pair of drawings and the amount of time it took subjects to make a positive match:

> Indeed this linearity (was) consistently found even in those subsets of the data that were obtained just a) for those pairs of objects differing by a rotation in depth or by a rotation in the picture plane, b) for the conditions in which the axis of rotation was held constant or was unpredictable, c) for each of the eight subjects individually and d), for each of the differently shaped isomeric pairs of three dimensional objects.[16]

And so was the 'Mental Rotation' experiment born. These relationships and contingencies were interpreted as manifestations of a single phenomena and hence explained by giving credence to the subjects reports that in order to perform the required comparison they first had to imagine one object rotated into co-incidence with the other, as if mental movies were being projected onto an inner screen for the perusal of a 'mind's eye' that has no prior knowledge of what the film is about and so must wait for events to unfold. If the 'speed' with which the image can be rotated in or on the inner screen is constant Shepard and Metzler's results seem to accord well with this common-sense proto-model. If it takes, say, half a second to verify that two drawings are in fact of the same object when they are rotated 30 degrees with respect to each other, and one second to verify that the same two depictions are identical when they are rotated 60 degrees with respect to each other, no other explanation seems motivated. No such result could be predicted a priori from the supposition that information about the shape and orientation of objects in the visual field is encoded in a discrete propositional or digital format. If the objects matched by the cognitive system were lists of sentence-like things - propositions - which describe (in some sense) the position and shape of the line drawings we could perhaps anticipate an effect upon reaction time due to object complexity (number of vertices etc.) but not orientation; for, in any likely algebraic model all rotations of co-ordinates of less than 360 degrees are of equal complexity. It doesn't take twice as long to multiply 5 by 4 as it does to multiply 5 by 2.

---

[16] Shepard and Cooper, *op. cit*, p. 43.

With this theme a number of theorists concurred, foremost of whom in terms of experimental output and theory construction has been Stephen Kosslyn and his co-workers. Within a few years Shepard and Metzler's methodology had been expanded and generalised to include a wide variety of putative imagery phenomena. In almost all cases highly significant relationships pertained between reaction time and subjective 'distance' within or between mental representations of visual objects, replicating and confirming the earlier results. As the erstwhile Myth of Mental Imagery became respectable in the eyes of modern Psychology, so a theory was required to make concrete the two distinguishable facets of the picture that seemed to be emerging from folklore and merging with Science: In the first case something manifestly spatially extended and picture-like seemed to be involved, thus requiring a non-circular formalisation of *pictorial* from any proposed theory; in the second case and not unrelated to the first, the picture-like representation is epistemically opaque to the rest of the cognitive system, otherwise no rotations or scannings of images would be required.

Kosslyn's model, which we shall take as representative of its school, sticks fairly close to the common-sense idea of an inner screen. Taking as inspiration the spatially extended display of a cathode-ray tube which may be generated from abstract, non-pictorial, descriptions in computer memory, spatiality is maintained by supposing that what we call 'images' are:

> temporary spatial displays in active memory that are generated from more abstract representations in long term memory ....

Such displays remain epistemically opaque as they require interpretation by a different part of the system prior to information being made available to that part of the system:

> Interpretative mechanisms work over ... and classify ... in terms of semantic categories ... as would be involved in realising that a particular spatial configuration corresponds to a dog's ear, for example....[17]

---

[17] Kosslyn et al, *op. cit.*, p. 536.

Stipulating that imagery occurs in the form of 'spatial displays' is not however sufficient to comprise an unequivocal theoretical assertion, for we still don't know what 'spatially extended' means apart from 'picture-like', and vice versa. Would not a propositional representation in the memory of a computer be spatially extended? Or are we to suppose that it would take up no space at all? But then how is a 'spatial display' different from a proposition? The issue thus devolves upon what it is exactly that spatial display cum imageshave that other forms of representation do not, an investigation which clarifies and defines the analog position in the course of justifying it.[18]

The qualities peculiar to images according to Kosslyn are their 'privileged properties'.It is these privileged properties that are the locus of disagreement as we shall see below. Certain other processes constitutative of the 'mind's eye' part of Kosslyn's model which work over the spatial display are from the perspective of critics of this position innocuous. It is the very idea of *picture-like* that has traditionally raised hackles, and continues to do so today.

These then are the privileged properties of imagery; the spatial medium, abstract spatial isomorphism, and abstract surface-property isomorphism.

The Spatial Medium.

Images are not things in themselves. Just as a painting is an arrangement of paint on a canvas so images must be supported by some sort of structure:

> Images occur in a spatial medium that is functionally equivalent to a (perhaps Euclidean) co-ordinate space ... locations are accessed such that the interval properties of physical space are preserved in at least two dimensions.[19]

Thus whilst image-representations may not themselves have distance and locations within themselves, they preserve distance and location information (in an unspecified manner) just as a Lands and Survey map

---

[18] I will commonly refer to analog processes as "images". This is for ease of expression and is not meant to beg any questions.

[19] Kosslyn (1980), p. 33.

will preserve elevation information without itself possessing a range of elevations. Kosslyn gives the example of a two-dimensional array stored in a computers memory; the relevant hardware will not arrange itself so as to possess a visible resemblance to a particular scene, should anyone prise the cover off and examine the area with a magnifying glass, but each part of the memory matrix corresponds to a spatially defined part, - or portion, - of the depicted scene, preserving geometric properties such as distance in a fashion *analogous* to the properties themselves.

Studies such as Cooper (1975) is typical of studies cited as illustrating this property. Cooper's subjects learned to distinguish standard versions of Attaneave polygons from reflected or mirror image versions at a single standard training orientation for each of the eight polygons used. Subsequently during the test sessions the same eight forms were displayed tachistiscopically at orientations differing by 0, 60, 120, 180 and 240 degrees from the training orientation. The subjects task was to identify as quickly as possible the target stimuli as either a standard or reflected form.

Over the course of 620 test trials a highly significant linear relationship was found to pertain between reaction time and angular departure from training orientation for both matches and mismatches - the classical "mental rotation" effect. Results were significant for all subjects and all orientations. No second order effects were found between figural complexity and the slopes and intercepts of the reaction time functions. This latter finding implies that images are rotated at a constant rate irrespective of their figural complexity, which in turn suggests that they are contained within a medium which has the power to move or rotate it's contents as a whole. To use an analogy, the locus of encoding (what is encoded) appears to be not the words but the page upon they are writ. A mode of representation which proceeds by encoding elements of the drawings, - vertices, segments etc. - could not be expected to be unaffected by the complexity of the figures encoded.

Again, the linearity of the reaction time function implies that the medium duplicates the isotropic interval qualities of Euclidean space. Furthermore, and most obviously, detailed information on the orientation of objects is preserved by whatever representation is being used. Such information is intrinsic to the nature of spatial representations, but must take the form of a postulated adjunct to more abstract forms of representation.

Abstract Spatial Isomorphism.

> Images are patterns formed by altering the state of local regions in the internal spatial medium. The pattern formed in the spatial medium is a topographic mapping from the represented object such that a) each local portion (set of contiguous points) of the image corresponds to a portion of the represented object as seen from a particular point of view, and b) the interval relations among the portions of the image implicitly represent ... the interval distances among the corresponding portions of the represented object. This property (is) 'abstract spatial isomorphism.'[20]

Not only is the medium in which images occur hypothesized to possess certain geometrical properties, so are the images themselves. As it seems unlikely that mental images are literally constituted by two dimensional slices of the primary visual cortex we need the next best thing, what a crumpled map or a map wrapped around an unevenly shaped object is to the same map spread flat upon a table. What the grooves in a record are to the sound they can produce (Sellars (1963)).

Here is one of the well known experiments that prompted Kosslyn to draw these conclusions:

Subjects were shown a drawing of a fictitious island which depicted a number of simple features such as a well, a hut, a tree and so forth.[21] They were then required to commit the map to memory in accord with a criterion of accurate reproduction of features to within 0.25 inches of their true location on a blank outline drawing of the island. Upon satisfying this criteria subjects were required to 'imagine' the map and 'mentally stare' at a named location. A word would then be presented which either did or did not name another location on the map. The subject was asked to scan across the image to the named location, if it was present, by imagining a flying black dot travelling from the one point to the other.

Reaction time for 'true' responses was found to increase linearly with respect to 'distance' scanned on the image. No such linear relationship was found in a subsequent study in which subjects were not explicitly instructed to use imagery. If this study is accepted uncritically it seems

---

[20] Kosslyn, *op. cit.*, p. 33.

[21] Kosslyn, Ball and Reisser (1978), experiment 2.

undeniable that the spatial metric of the map has been transferred to the internal representation in the course of memorization. This would appear to be the implication of the correlation found between reaction time and the distance between two points on the map, irrespective of what the two points may be and which particular portion of the map is transversed. For instance two points adjacent on the map are 'adjacent' but distinct on the image. Without being committed to supposing that the image is laid out in an island shape in the brain, it is clear that the interval relations of the map, which are themselves ultimately only definable in terms of relationships between velocity and time (a result of the theory of relativity), are recreated in terms of relationships between a constant 'scanning velocity' and time in the representation. The existence of a 'scanning time' also relates to the epistemic opacity we mentioned above; the location of all things on the island is not known at once. Position makes a difference to response latency, which seems explicable only if ones attention or 'mind's eye' has to move to a certain feature on the representation in order to make its nature available to the rest of the cognitive system. Access to an answer does not seem to be limited in this way when we consider purely abstract questions like 'what is 3 times 12 ?' and 'what is the capital of Zambia ?'.

Abstract surface-property isomorphism.

Images are also held to preserve information about the secondary qualities of the objects depicted, further emplacing the intuitive 'picture in the head' model:

> Images ... depict information about the appearance of surface properties of objects, such as texture and colour ... Thus, although the image itself is not mottled, or green, or bright , or faded, it can attain states that are interpreted as evincing these properties.[22]

So, as Kosslyn puts it, "a portion of an image of a green thing ... is an image of a portion of a green thing".[23] Possessing an obviously lower priority than the demonstration of the existence of representations with

---

[22] Kosslyn, *op. cit.*, p. 34.
[23] *ibid.*

properties analogous to spatial extension - hence the term *analog spatial medium* -, the literature directed at demonstrating the presence of this property in particular has not been copious. It is, for instance, difficult to disassociate colour and texture from extension thus the demonstration of just the former is tricky, and perhaps not necessary if the latter has already been shown. Nevertheless Finke and Schmidt (1977) managed to elicit the McCollough effect of orientation specific coloured after-effects - such as are ordinarily due to exposure to a brightly coloured patterned stimulus - by means of a condition in which subjects were required only to imagine a certain colour on a black and white patterned background. They concluded:

> The present study demonstrates that weak feature-contingent colour aftereffects may be produced following adaptation procedures in which imagination instructions are substituted for the physical presence of colour and pattern stimuli.[24]

Because the test for the presence of the McCollough effect involves a forced-choice discrimination procedure for which the effect-revealing responses are not intuitively obvious, the results obtained are well controlled for task demand biases. The clear implication is that colour information may be accessed from memory and 'displayed' at a stage which may share many of its functions with perception, as would be expected upon the supposition that imagery functions allow the perusal of an imaginal display in much the same fashion as an actual visual scene is examined.

The Model

The studies described above are representative of a substantial body of literature which may be interpreted as supporting the existence in cognition of a class of representations possessed of the privileged properties typical of what we would intuitively call 'mental imagery'.

A model devised to account for these findings has been described in Kosslyn (1980), (1981), and Kosslyn et al (1979) and it is to this tradition that we will refer. Shepard and associates have developed a similar model (cf Shepard (1984)), but the issues remain fundamentally the same, so for

---

[24] Finke and Schmidt, (1977), p.

simplicity we will confine ourselves to the Kosslyn tradition, which today remains the most explicit and well described analog theory of imagery.

Kosslyn has closely followed the intuitive distinction between images themselves and inspections of them in the construction of an explicit computer model of imagery and mental scanning processes, hence a division of labour is incurred in the model between media and data structures on one hand, and processes operating upon these on the other.

In the first category we have the medium of image representation, the 'spatial medium' or 'visual buffer', which is a kind of operationally defined or functional space where the operations concerned come under the headings 'formatting' and 'accessibility'. The format of a medium places restrictions on what sorts of data structures can be encoded within it, and its accessibility is defined in terms of what processes can access data structures within the medium. The format of the visual buffer, naturally, is appropriate for the support of 'surface images'.

In the second category the images or 'surface displays' or 'surface representations' are data structures which "depict" information wherein "every portion of the representation must correspond to a portion of the object ..."[25], in other words the image cum data structure possesses 'abstract spatial isomorphism'. Images are depicted in the visual buffer by virtue of being 'patterns of activation' in the medium itself, and this is mimicked in a computer simulation by the selective filling in of cells in a matrix data structure. Data structures of this form whilst not literally laid out in a two dimensional plane like a picture preserve the functional inter-relationships between their 'regions' in the manner of their computational formatting and accessibility.

In a sense the image medium is defined by limitations upon its accessibility so that, for instance, two patterns of activation cannot be identified as the same even if they are until a process called "Rotate" has operated upon the two so as to bring them into co-incidence. This in turn is necessitated by the inability of inspection procedures with the intuitively perspicacious names of 'Lookfor' and 'Find' to match objects unless their orientations are identical. In this fashion the basis of spatiality and image distance effects in the model is the permitted orderings of the processes acting upon the data structures and the degrees of freedom in the data structures themselves, which are not un-coincidentally the same degrees of freedom possessed by locations in a two dimensional plane.

---

[25] Kosslyn, (1981), p. 50.

The information to be displayed in the spatial medium is stored in list structures in memory in the computer simulation and accessed by name. Literal (appearance) and propositional information is stored such that the model can generate images of objects by name and answer questions about their constituent parts. Insofar as this part of the model is not supposed to manifest any of the privileged properties hypothesized as being characteristic of image representations, their format and structure is theory neutral.

In the third category we have those processes which operate on the medium. These are interdefined with each other and by their access to the medium. For instance LOOKFOR is an executive routine which calls other processes such as ZOOM, REGENERATE, FIND etc. Each of these is a function which takes as input a set of co-ordinates or the name of an object or part of an object and produces as output either either an affirmative or negative response, or an alteration in the pattern of activation in the matrix data structure. There are numerous processes in the model of image generation, inspection, transformation and classification, each performing operations on image or memory structures aptly described by their names viz., PICTURE, PUT, PAN, SCAN, etc.

Most of the explanatory work is done by three properties natural to the model. Firstly; it is postulated that image transformations take place incrementally, such that the time required will be proportional to the degree of the transformation. This accounts for a large number of results concerning image rotation and mental scanning phenomena. Incremental transformation is principled within the model because it insures that the target is not overshot, that the resolution and point of focus are appropriate for comparison, and that the image does not 'break up' as it might do if subjected to a single large transformation.[26] Secondly; the medium is postulated to have a limited resolution, limited extent and an area of optimum resolution. In this manner the data concerning the necessity of imagining objects at a given size - not too large and not too small - is encompassed, as well as the mental scanning data wherein subjects apparently have to move their focus of attention in order to 'see' the different parts of a larger image.[27] Thirdly; and in general, tasks will take longer to complete the more subcomponents have to be called. Thus it takes more time to inspect and generate an image with more nameable

---

[26] Kosslyn et al, *op. cit.*, p. 542.
[27] cf Kosslyn, (1975), (1976).

parts, more time to find rotate and match than just to rotate and match etc. [28]

With a great deal of interdefenition of component parts Kosslyn's model opens itself up to accusations of circularity. This is of the nature of functional and operational definitions, wherein one component or quality is defined in terms of the effects of something else upon it. For instance, the spatial medium is partially defined by its format, which is the format appropriate for supporting quasi pictorial data-structures, however "The surface representation is a quasi-pictorial representation that occurs in a spatial medium"[29]. If the 'analog medium' is circumscribed only by its interrelationships with the processes that operate upon it then there would seem to be little more to the model than a sort of abstract calculus with suggestive names attached to different parts of it. What claim, if any, is being made about image representation if to be a representation of a given format is simply to be accessed in a certain way ? The gist of the analog position so far seems to be that the representations we call 'images' behave *as if* they are spatially extended, but this is a) trivially true, and b) exactly the claim of the opponents of this view of imagery. Pylyshn claims, for instance, that we have *tacit* knowledge of the geometrical qualities of physical objects, which is however encoded in in a propositional or sentence-like format (see below, Ch. II). Given the above evidence,how can anyone deny that the representations involved in a certain kind of task, in the main, behave in a fashion *analogous* to an extended two-dimensional plane? Thus if the analog theory is to be rescued from vacuity it is important, I think, that the interpretation of function implicit in functional definitions of the spatial medium is directed towards the functional properties of the brain as opposed to a response directed, input-output sense of function.

An analogy is to think of a black box with two slots on opposite sides. If a blank piece of paper is placed into slot A, after a short period of time and a few clicks and whirs a beautifully hand written letter of condolence will emerge from slot B. Now, to merely say that the box 'writes a letter' is to say nothing about the internal constitution of the box. Whilst 'letter writing' describes its function from outside, the letters themselves could perhaps be cleverly printed or faxed from elsewhere. If a theory of the box is to make substantive claims about its internal mechanisms it must

---

[28] cf Kosslyn et al, *op. cit.*, p 542.
[29] Kosslyn (1981), p.49.

leaven its functional descriptions with some degree of mechanical (in a general sense) meaning, for instance; '... the computing unit chooses from amongst the handwriting styles of Jane Austen, Emily Bronte etc. and sends messages to the mechanical hand ...' or '... the box contains a man with a telephone who is paid $20 an hour ...'.

Thus whilst Kosslyn and his co-workers are frequently equivocal as to the sense of function, analogy and 'as-if' they intend, the theory relevant sense must be taken as given in Kosslyn's discussions of format and 'the functional capacities of the *brain* '. Here Kosslyn makes use of Nelson Goodman's theory of notation in order to remark that the imagery issue is a dispute as to the format of 'image' representation. The format of a code or type of notation according to Goodman concerns:

> the nature of the inscriptions used ... and how these inscriptions are mapped onto (used to represent) compliance classes.[30]

'Inscriptions' are not tautologically described in terms of the functions of other things because not all inscriptions are members of a character class, consequently not all inscriptions have a function. Certain inscriptions are members of the same character class as 'b', for instance, but shapeless squiggles are not members of any class. If an inscription is an ink-mark then not every ink-mark will be a letter. [31] Keeping with this analogy, image "patterns of activation" are like inscriptions. They are not things which merely behave as-if they are spatially extended any more than some inscriptions merely behave as-if they are characters. In each case some part of their definition is intrinsic and non - functional.

This is how Kosslyn sees his theory:

> A cognitive account of imagery is a theory about the functional capacities of the *brain* .[32]

> 'Mental events' (are) events described at the level of functional states of the *brain* . These states arise from the brain but are not necessarily identical to particular

---

[30] Kosslyn (1980), p.30.
[31] Goodman (1969), Ch. 4.
[32] Kosslyn (1981), p.47.

> configurations of neural activity ... it may well be that the only thing the different neural states that correspond to a 'functional brain state' have in common is their role in promoting a particular kind of computation ...[33]

A computer program is:

> a description, at a particular level of analysis, of what the *machine* will do given specific inputs.[34]

In each case the abstraction involved in identifying something by its function lies in the mode of description, not the thing described. It is clear from the above quotes that Kosslyn intends his theory to take the form of a generalized description of operations of the brain and neurological structures. Similarly, his concept of computer program is one wherein a computation is a generalized description of the operations of a mechanical object. In light of these considerations it is clear that the analog spatial medium is not merely something that occurs in between LOOKFOR and ROTATE which mimics (in some unspecified way) the behaviour of a representational structure which is in some manner extended; - it actually is, according to the analog theory, an extended neurophysiological structure which befits the display of visual information by being continuous, isotropic, and a fairly simple topological deformation of a two-dimensional plane. If Kosslyn is saying anything at all he is saying something almost physiological. His theory is like a simplified diagram of the actual physiological processes involved, because we simply don't know enough about brain physiology to put the hypothesis any more precisely. The diagram will be somewhat abstract and function-based, but the things diagrammed are quite concrete. (Theoretical models delimit a range into which the properties of the process modelled will fall. Theories never say that the subjects of the theory are *exactly* like the model, cf Sellars (1963) ch. 4, Nagel (1961)).

These considerations are important in order to combat any suggestion that the analog medium is an abstraction, in which case the actual stuff which does the 'rotation' calculations could be a fine network of propositions which just behave as-if they are continuous and extended.

---

[33] Kosslyn (1980), p114.
[34] ibid.

Whilst we have not discussed exactly what the neurological basis of a sentence of mentalese, a list structure or a proposition might be like, they would certainly be nothing like the structures implied by the above model.

Many criticisms have been levelled at this theory; I intend to completely ignore all but one, which is sketched below. Pylyshn's Tacit Knowledge argument, which makes significant use of his notion of 'cognitive penetrability', is I feel the only argument which seriously threatens the analog hypothesis, and is the only argument of anything but local interest. I will not defend this contention here for it would take us too far afield. My argument below does not depend upon Pylyshn's thesis being demonstrably the 'most important' criticism of the analog theory. However, I show that this criticism is in fact *fatal* to the analog position unless shown to be invalid, and you can't get much more important than that. For a general review of the spectrum of arguments that have been levelled against imagery in recent years see Kosslyn (1980).

COGNITIVE PENETRABILITY AND TACIT KNOWLEDGE.

The most significant threat to a theory which treats images as functional representational formats and cognate to a particular 'capacity of thebrain'goes something like this: These are only two ways in which a given cognitive act or mental performance might be explained. The first type of explanation, the 'knowledge based', 'tacit knowledge', 'semantic level', 'symbolic' or 'cognitivist' account appeals to:

> symbolically encoded facts about the world and to rules for transforming representations and drawing inferences.[35]

Cognitive tasks are performed by symbolic representations and rules for operating upon these, just as computers transform symbolic expressions according to the rules of a program.

The tactic of positing representations in psychological explanation arises as a direct result of the plasticity of intelligent behaviour with respect to the environment. Human behaviour is in crucial respects not mediated by the objective features of the world, but by the way the world is

---

[35] see Pylyshn (1978), (1979a), (1979b), (1980), (1981), (1984).

'taken' or 'represented' by the individual. Thus someone might shout 'Fire ! Fire !' and leave her office in a hurry even if there is no fire, merely a cigarette smouldering in a waste paper basket. And this one piece of information may be conveyed by a variety of objectively different features of the environment, such as a phone call, an acrid smell, or a crackling sound coming from the next room, all of which have nothing by way of 'stimulus properties' in common. In short, intelligent behaviour is always mediated rationally on the basis of what a person knows, or thinks she knows, about the world, and by the meanings (roughly speaking) of external events.

A further important consideration is that although it is trivially true that all thinking is done by and in the brain, the laws governing relations between representations are not reducible to laws governing physiological structures. 'Thoughts' may be defined functionally by their role in the cognitive economy, but just as a transfer of funds from A to B may be effected in a variety of physically different ways - by coin, cash, cheque, IOU or telegraphic transfer - and still be the same economic event, so two instances of a given representation governed cogitation may have nothing in common physically. In other words there will be token but not type reduction (cf Fodor (1975) introduction).

Of course as materialists we suppose that representational thought is governed by the structure of the brain somehow, but by virtue of being the basic level operating principle of cognition - that which explains the basic links or syntax of symbolic inter-relationships - and by virtue of being physical, the hardware or 'functional architecture ' of the brain is fixed with respect to representation governed processes. This brings us to the second type of explanation, the 'functional architecture' or 'syntactic' account, which proceeds by appeal to:

> (the) intrinsic lawful relations among properties of a particular physical instantiation of a process.[36]

If we accept this then it is to be expected that the fixed functional properties of the brain will not be susceptible to influence by the meaning of intellectual tasks, or by purely cognitive variables such as beliefs and goals. This provides us with a decision procedure for telling which processes are in fact instantiated in the functional architecture viz.:

---

[36] Pylyshn (1981), p.19.

> A function that is alterable ... in a way that is rationally connected with the meaning of ... inputs ... is said to be cognitively penetrable.[37]

Now, all of the transformations imputed to 'images' in the 'spatial medium' would appear to follow rationally from their initial states. For instance, that it takes longer to rotate an object through 180 degrees than 120 degrees, elephants are larger than flys and so easier to see at a distance, it takes longer to traverse larger distances etc.

A process which can be so influenced by beliefs and the semantic interpretation of whatever is the object of the process - in the present case physical objects and their transformations - will then be properly explained by appeal to representations and symbolic processes. The process itself will contain at least some representation or rule governed components, and so cannot be instantiated as a functional capacity of the brain, as Kosslyn's spatial medium is supposed to be.

The moral then for a theory whose central claim is that images are functional representations which possess spatial properties themselves is that it has failed to distinguish between two domains, or two types of principal; the extrinsic semantic domain, and the intrinsic syntactic domain. Thus, what is allegedly confused are the properties of represented objects, - in particular their spatial extent, - and the properties of representations themselves. To do this, it is alleged, is to commit a kind of category mistake;

> This tendency ... leads to a way of stating the principles by which mental processes operate that deprives the principles of any explanatory value, because it involves principles expressed in terms of properties of the represented object rather than in terms of the representations structure or form. Yet expressing a principle in terms of properties of the domain represented begs the question why processing occurs in this fashion. The mechanism has no access to properties of the represented domain except insofar as they are encoded in the representation itself. Thus a principle of mental processing must be stated in terms of the formal structural

---

[37] *ibid* , p.20.

> properties of its representations, not what they are taken to represent in the theory.[38]

In other words, generalisations expressed in terms of concepts attributable to the objects of the external semantic domain, such as shape and orientation, are properly referred to the first kind of explanation we encountered above, the 'knowledge based account'. By virtue of this fact they are not to be accounted for in terms of the second type of explanation encountered, the 'functional architecture' account, empirical results notwithstanding. Given that the analog theory seeks to interpret a certain body of empirical evidence in terms of the "intrinsic properties of the representational medium",[39] it is an explanation of this latter type, and is thus invalid, given what it seeks to explain.

[38] Pylyshn (1984) p.230.
[39] *ibid.* p.233.

II THE THEORETICAL BASIS OF COGNITIVE PENETRABILITY.

In order to effectively analyse the import of Pylyshn's argument against the Analog position we must first break it down into its constituent stages. Once his basic premises have been isolated and clarified it will be possible to assess whether they are well founded and whether his conclusions indeed follow.

Such delving into the structure of arguments is not always to the point in Science, where it is to be expected that most questions are empirical or evidential. However, the present issue depends upon the interpretation of an extant, more or less complete and unchallenged empirical literature. This has been explicitly recognised by Richardson:

> Certain basic facts are not in dispute. For example, when subjects are instructed to perform a mental scanning task response times are found to be analogous to those obtained when performing the same task physically, in the external world. Objects, whether manually or physically represented, take more time to locate when they are more distant from each other. What is in dispute is the process postulated to account for these results.[1]

Pylyshn's argument contains certain weaknesses in its use of the Cognitive Penetrability criterion as a theoretical principle for the judgement of imagery and related tasks. The rationale of this principle is such that it in turn depends upon some quite basic suppositions about the nature of cognitive explanation. It is to these concerns that we will now turn, beginning with the implications of Pylyshn's disavowal of any form of behavioural reduction.

---

[1] Richardson (1979), p.563.

TWO KINDS OF BEHAVIOURISM.

Like Jerry Fodor before him (in his 1975 'The Language of Thought') Pylyshn finds that one of the first concerns of a sympathetic analysis of the foundations of Cognitive Psychology is to defend the integrity of the "explanatory vocabulary of cognition"[2] - the intentional vocabulary of beliefs, desires, inferences, goals etc. - against behavioural and physiological reductions which would deny the scientific validity of these mentalistic terms.

The first of these reductions comes under the rubric of Behaviourism, and may be divided into two strands of argument; philosophical or 'logical' Behaviourism, and methodological arguments for Behaviourism.[3]

The first is characterised by a tendency due to the Ryle/Wittgenstein school of philosophical analysis to simply deny that there are such things as 'mental happenings' on the grounds that mental terms are no more than a sort of shorthand for varieties of behaviour:

> It is being maintained ... that when we characterise people by mental predicates, we are not making untestable inferences to any ghostly processes occurring in streams of consciousness we are debarred from visiting; we are describing the ways in which these people conduct parts of their predominantly public behaviour.[4]

We can see from this that Ryle's argument has to do with the alleged conditions for the justified appellation of mental terms, as the title of his magnum opus 'The Concept of Mind' implies. We can form a working definition of this form of Behaviourism after Fodor;

> To qualify as a Behaviourist ... one need only believe that the following proposition expresses a necessary truth; For each mental predicate that can be employed in a psychological

---

[2] cf Pylyshn (1984), ch. 1.
[3] For this dichotomy see Shaffer (1968), p15.
[4] Ryle (1949), p.51.

> explanation, there must be at least one description of behaviour to which it bears a logical connection.[5]

'Logical connection' here means that at minimum it is sometimes possible to deduce the truths of mental ascriptions from the truth of behavioural ascriptions.[6] As a logical connection is necessary, as opposed to contingent, it follows that mental states are logical constructions of behaviour:

> For those influenced by the tradition of logical behaviourism, (mental) phenomena are allowed no ontological status distinct from the behavioural events that psychological theories explain. Psychology is thus deprived of its theoretical terms except where these can be construed as nonce locutions for which behavioural reductions will eventually be provided. To all intents and purposes this means that psychologists can provide methodologically reputable accounts only of such aspects of behaviour as are the effects of environmental variables.[7]

The error behind the 'Cartesian Myth' of inner psychological events is said to be that of a 'category mistake' which:

> represents the facts of mental life as if they belonged to one logical type or category ... when actually they belong to another.[8]

Just as we may make the mistake of assuming that there is some single thing which is the University of Canturbury over and above the sum of its staff, students, buildings etc., there is supposed to have been ingrained both into philosophical and common knowledge a taking of members of the categories of dispositions (to behave) to be of the category of occurrences, which are of a different logical type. For instance, we take the attribution of a particular motive to someone to be an assignment or

---

[5] Fodor (1968), p.51.
[6] cf *ibid.*, p.56.
[7] Fodor (1975) p.1.
[8] Ryle, *ibid*, p17.

identification of a particular something 'in his head', whereas according to Ryle all we are doing is describing "the sorts of things he tends to try to do".[9]

In a similar vein Wittgenstein pronounces that:

> an inner process stands in need of an outward criteria.[10]

This is not intended as a simple epistemological point. The significance of behavioural criteria for psychological concepts is that we allegedly cannot mean a), just one process and b), an 'inner' process by a mental word such as 'understand'. The invisible something which is held in common by all examples of understanding behaviours, but which can never be detected over and above the detection of these behaviours, cannot be referred to except by referring to these behaviours, so we cannot say anything *about* it. If we can say nothing about a process then we cannot say that it is the same process on occasions a, b and c, - indeed, we cannot even say that it is a *process,* for there for there may be no one thing or essence in common on all three occasions. Wittgenstein's famous example is 'game'; there is no one thing that all things we call a 'game' have in common. They don't all have winners and losers (solitaire), teams (noughts and crosses), rules (a child playing), or even human participants (two computers playing chess against each other). There is nothing that it is necessary for every game to have, they have, rather, a *family* of resemblances: "phenomena have no one thing in common which *makes* us use the same word for all".[11]Elsewhere he writes:

> Try not to think of 'understanding' as a 'mental process' at all. For *that* is the expression which confuses you. But ask yourself: in what sort of case, in what kind of circumstances, do we say 'Now I know how to go on'.[12]

In light of this Wittgenstein's attitude to mental processes can be aptly summarised by his epithet:

---

[9] *ibid,* p.112.
[10] Wittgenstein (1953 , section 580.
[11] *ibid* , section 65, My emphasis. For games see section 66.
[12] *ibid* , section 154.

> a nothing will do as well as a something about which nothing can be said.[13]

These denials of existence can be distinguished from the methodological themes which have been prevalent in psychology for much of this century.

As a reaction against the ascendant methodology of the time, Introspectionism, Watson in his landmark 1913 paper defined Psychology as:

> a purely objective experimental branch of natural science (the goal of which is) the prediction and control of behaviour.[14]

Consequently, amongst other things:

> the time honoured relics of philosophical speculation need trouble the student of behaviour as little as they trouble the student of Physics.[15]

To the same ends B.F. Skinner was influenced in his early years by Bridgeman's Operationalism,[16] a doctrine which sought to embed the validity of scientific concepts in their relationship to public operations. For S.S. Stevens, for instance:

> Science ... is a set of empirical propositions agreed on by members of society ... Only those propositions based on operations which are public and repeatable are admitted to the body of Science.[17]

At this time the philosophy of science was dominated by a cluster of 'isms'; Logical Empiricism, Positivism, Instrumentalism and Operationalism, which held in common three basic themes:

a) there is a theory-neutral 'observation language' which is the incorrigible basis of all human knowledge.

---

[13] *ibid* , section 304.
[14] Watson (1913), p.158.
[15] *ibid* , p.166.
[16] see Skinner (1931), (1979), p.41.
[17] Stevens (1939), quoted in C.E. Buxton (ed.), (1969).

b) scientific knowledge, being empirical, testable and so on, must then be *about* these observables.

c) talk of unobserved theoretical entities is only meaningful insofar as these entities can be translated or reduced into observation terms, or the public and repeatable operations of scientific intruments. These sorts of attitudes lead to statements like:

> Science is ultimately intended to systematise the data of our experience.[18]

and;

> A scientific theory ... tells us no more than it appears to tell us about the experimental facts, namely that they may be related in a particular manner.[19]

To this mood can be added two sound methodological principles;

a) All things being equal, the simplest theory that accounts for the data should be preferred.

b) All things being equal, the weaker more testable and hence more falsifiable theory that accounts for the data should be preferred, (the empirical content of the theory should be maximised). Thus the more directly the theoretical terms and intervening variables are related to the data the better the theory.

If these points are coupled with a liberal dose of old fashioned empiricism the familiar dogmas of Behaviourism emerge; mental/cognitive (i.e. theoretical) terms should be eliminated, or are irrelevant; Psychology should concern itself only with the relationship between environmental factors and behaviour (i.e. observables). Accordingly in his paper 'Are Theories Really Necessary ?' Skinner hopes that even higher mental processes can be accounted for atheoretically:

> (they) ... appear to be susceptible to formulation in terms of differentiation of concurrent responses, the discrimination of stimuli, the establishment of various sequences of responses and so on. There seems to be no a priori reason why a

---

[18] Hempel (1959), quoted in Feyerabend (1981), p.17.
[19] Crombie (1953), quoted in Pears (ed.), (1962), p.69.

> complete account is not possible without appeal to theoretical processes ...[20]

It is against this background that Pylyshn presents his case. Whilst Behaviourism is not a predominant influence nowadays the Behaviourist's arguments are the first fence to be crossed by any attempt to establish cognitive 'mentalistic' explanation from first principles.

His tactic is to argue obliquely that a behavioural taxonomy of Psychology would fail to capture a class of generalisations, couched in the terms of a pre-existing cognitive vocabulary, which seem to constitute a domain appropriate for study irrespective of Behaviourism's greater methodological propriety:

> the problem in explanation in Psychology forces us to adopt a certain kind of taxonomy of behaviour. Specifically, it is because explanations attempt to capture generalisations that we find ourselves forced to resort to what I call a cognitive vocabulary in revealing certain fundamental patterns of intelligent, largely rational behaviour.[21]

This point is conveyed by means of an example:

> Suppose you are standing on a street corner and observe a sequence of events that might be described as follows; A pedestrian is walking along a sidewalk. Suddenly the pedestrian turns and starts to cross the street. At the same time, a car is travelling rapidly down the street towards the pedestrian. The driver of the car applies the brakes. The car skids and swerves over to the side of the road, hitting a pole. The pedestrian hesitates, then goes over and looks inside the car on the drivers side. He runs to a telephone booth at the corner and dials the numbers 9 and 1.[22]

Question: Why did the pedestrian go to the phone booth ?

---

[20] Skinner (1950), p.215.
[21] Pylyshn *op cit.* p.2.
[22] *ibid*, p.3.

Presumably there will exist in principle a true explanation of these events in terms of Physics, Chemistry, and/or an objective description of behaviour. However, notwithstanding the enormous practical difficulties inherent in any account of human action in these terms (sheer length being predominant amongst these), it is also true that explanation is relative to the explainer's interest, and to the sorts of questions asked before the investigation has begun. Thus truth is not a sufficient condition for explanation. Pylyshn conveys this idea in terms of the referential opacity of the word 'explains'.[23] Roughly speaking, the modus operandi of deductive nomological explanation - the classical interpretation of theoretical explanation - is the subsumption of the 'facts' under a natural law or lawlike generalisation. But there are no a priori limitations on what can count as the initial fact, and you can generalise over anything that occurs more than once. Hence Pylyshn argues that as long as we are interested in the whys, whats and wherefores of rational human behaviour , explanation must use the intentional terms which are the initial explananda, i.e. belief, memory, knowledge, desire, recognition etc. It is only when described in these terms that regularities emerge to be generalised about.

A thoroughly objective behavioural account of the above event would presumably amount to an exhaustive catalogue of movements in the presence of objects, down to the number of footsteps taken and which finger is used to dial, for this would seem to be the only level of description which is appropriately free of mentions of aims, goals, intentions and the like. This runs into difficulties when we consider that there may, for all we know, be 2000 ways of taking a step or dialling a phone number. What laws could the very unique and rare sequence that actually occurred be subsumed under ? Yet upon what grounds can a Behaviourist group together instances of behaviour in order to have something to generalise about without hiding teliological or intentional notions in the new description ? A telephone for instance is essentially a device with a certain use, - communication, - thus it has scant a priori bounds upon its physical structure. How then can behavioural interactions with telephones have anything in common without some notion of use, purpose or intention lingering in their description ? The cognitive vocabulary systematises an otherwise massively complex world.

---

[23] *ibid*, p.4.

> If a person *knows* how to get out of a building, and *believes* the building to be on fire, then generally he will set himself the goal of being out of the building, and use his knowledge to determine a series of actions to satisfy this goal. The point is that even so simple a regularity could not be captured without descriptions that use ... mentalistic terms ... because there is an infinite variety of specific ways of 'knowing how to get out of the building' of coming to 'believe that the building is on fire' and of satisfying the goal of being out of the building. For each combination of these an entirely different causal chain would result if the situation were described in physical or strictly behavioural terms. Consequently the psychologically relevant generalisations would be lost in the diversity of possible causal connections.[24]

By this token the pedestrian example would be related to a series of generalisations that went something like 'Generally, if a person believes he/she has witnessed an accident, and infers that medical assistance has not already been summoned, he/she will desire to call for assistance ... ' and so on. Unlike any physical law that it may be possible to formulate, the cognitive generalisation will be counterfactual supporting; for instance, the exact same accident physically speaking would not inspire the pedestrian to phone for help if he believed that he was on a movie set and that it was a harmless stunt. Conversely, an alteration in the physically or behaviourally described sequence does not necessarily change the cognitive sequence; for instance, it makes no difference if the pedestrian walks, runs or crawls to the phone box. Hence beliefs and desires predict where behavioural and situational facts will not.

This is tied up with the stimulus independence of cognitive behaviour. It is granted nowadays that:

> virtually no candidate physical properties (for example particular physical features) are either necessary or sufficient for a person perceiving some situation in a certain way - for perceiving the stimulus *as* a something in the distal scene.[25]

---

[24] Pylyshn (1980), p.112.
[25] Pylyshn (1984), p.13.

To perceive-as is an epistemic state, thus we cannot be guarantied that our pedestrian will form a certain particular set of beliefs in a given situation on any two occasions, even if the stimulus conditions are exactly the same. The events may be 'taken' in different ways, to quote the adage, 'perception is theory laden', which is another way of saying that perception depends as much upon pre-existent beliefs as it does upon stimulus conditions.[26] To summarise:

> it is ... the environment or the antecedent event as seen or interpreted by the subject rather than as described by physics, that is the systematic determiner of actions, and ... actions performed with certain intentions rather than behaviours as described by an objective natural science like Physics, that enter into behavioural regularities.[27]

The same argument applies to the fact that the sequence of behaviours are in principle predictable from the laws of nuerophysiology coupled with knowledge of the pedestrian's initial nuerophysiological state and certain assumptions about physiology and anatomy. Despite this the actual sequence of neural events which led to the dialling of 9's and 1's in the above example are neither necessary nor sufficient for the act being an instance of dialling for help. In the first case we can imagine exactly the same sequence of events cognitively speaking being realised by a martian or an automaton with nothing physiologically in common with the pedestrian. In the second case we can imagine that the phone number of the emergency services is composed of 2's and 3's so that the given physiological sequence has a different effect. Consequently the physiologically described sequence is psychologically indeterminate.

It might be replied to this that whilst the establishment of a cognitive vocabulary for the making of useful generalisations may have force against methodological behaviourism, the utility of certain descriptive terms does not imply that the things that they refer to actually exist as a separate kind of thing. It could be held that intentional generalisations are a sort of useful category mistake; that what we really mean when we say

---

[26] cf Fodor and Pylyshn (1981).
[27] Pylyshn (1984), p.9.

that someone believes or intends such and such is that he/she is inclined to behave in a certain way (Ryle) does not deny that a large number of ways of behaving might be cogently summarised using the words 'believes', 'intends' etc.

That Pylyshn specifically disavows this interpretation is an important lemma in his argument. He is thus not merely saying that cognitive generalisations are useful and systematise an unruly domain, but also countenancing a class of entities to which the cognitive/intentional generalisations refer, and which are not even in principle reducible to facts about behaviour, the brain, or the environment, that is;

> the kind predicates of the special sciences cross-classify the physical natural kinds.[28]

For a superior level of explanation to reduce a subordinate level the types of the former must reduce to the types of the latter. The physiological reductionist for instance argues for the identity of neurological and cognitive event *types*. That every cognitive event token is a neurological event is uncontroversial; this is the supposition that all cognitive events are brain events. But only type identity ensures the reducibility of cognitive *laws* to neurological laws.[29] By the same token, Pylyshn is concerned to show that his cognitive level phenomena do not stand in a type-type relation with any other level of explanation:

> (the cognitive taxonomy) is not only more abstract than (Physics and Biology) it classifies events in equivalence classes whose boundaries typically do not coincide with the boundaries of classifications based on the other sciences. (In Psychology as in all the special sciences) in each case generalisations from each science are stateable only over their own special vocabulary; consequently, the lawlike generalisations of these sciences are not reducible to some finite combination of physical laws. Each category ... stands in a type token relation to categories of Physics, which means a

---

[28] *op. cit.*, p.18.
[29] cf Fodor (1975), ch. 1.

> category ... cannot be reduced to a finite disjunction of the categories of Physics.[30]

COMPUTATION AND COGNITION.

Given that people behave in accordance with their beliefs, desires, hopes, fears and so on, in what fashion can these entities enter into psychological generalisations when their objects may be non-existent, or not credibly in a causal relation with the psychological subject ? How can my behaviour be influenced by my beliefs about unicorns, future events, the north pole or Jupiter ? Certainly not by means of any causal relationship or disposition I may have towards these things. It becomes necessary to posit the existence of internal representational states:

> the causes of ... behaviour are not literally numbers, anticipated future events, or other 'intentional objects', but rather some physically instantiated internal representation of such things.[31]

> How can the state transitions (of a computer) depend both on physical laws and on the abstract properties of numbers ? The simple answer is that this happens because both numbers and rules relating numbers are represented in the machine as symbolic expressions and programs and it is the physical realisation of these representations that determines the machine's behaviour.[32]

There are two strands to this proposition, the first of which is the 'proprietry vocabulary hypothesis'. This is the idea that there is a unique and exclusive level of description appropriate for the expression and explanation of cognitive phenomena, and that the vocabulary of this level of description is that of computational alogarithms:

> the privileged vocabulary claim asserts that there is a natural and reasonably well defined domain of questions that can be

---

[30] Pylyshn (1984), p.21.
[31] *ibid*, p.26.
[32] Pylyshn (1980), p.113.

> answered solely by examining 1) a canonical description of an alogarithm (or a program in some suitable language...), and 2) a system of formal symbols (data structures, expressions), together with ... a 'regular scheme of interpretation' for interpreting these symbols as expressing the representational content of mental states (i.e., as expressing what the beliefs, goals, thoughts and the like are about, or what they represent).[33]

Succinctly, this is the domain of "cognitive rule governed processes acting on semantically interpreted representations".[34]

The second strand to the proposition is the tripartate organisation of explanatory levels which this necessitates. These levels are:

1) the biological or physical level.
2) the symbolic or functional level.
3) the semantic or intentional level.

Allegedly each level in ascending order instantiates and thus explains the basic working principles of the one above, whilst each level in descending order owes its existence to those generalisations which cannot be stated in or captured by the next level down. Pylyshn refers to this tri-level organisation as "the basic assumption of cognitive science ".[35]

Our first priority when confronted with a putative cognitive regularity is to explain it in the most parsimonious way possible. This is Occam's familiar methodological principle. Hence if a cognitive phenomena can be subsumed under physical or biological principles then we need go no further.[36] But as we have seen, a physical explanation of cognitive process is unlikely to pertain, in which case we must refer the explanation 'upwards' to the symbol level. At this level formal symbols are thought to be manipulated by alogarithmic or computational processes. Those cognitive generalisations which cannot be captured at this level must be referred to the semantic or intentional level, which is also the level of

---

[33] *ibid* , p.116.
[34] Pylyshn (1981), p.25.
[35] Pylyshn (1984), p.131.
[36] cf *ibid.*

ordinary language, folk psychological ascriptions of beliefs, desires, hopes and the like.[37] This is the 'instantiation hierarchy'.

Conversely in top-down mode, we proceed from the epistemically proximate, - the ubiquitous beliefs, desires and behaviours due to these which are our immediate 'things to be explained', - and seek to explain these regularities by exhibiting them as the product of symbol level computational processes. This tactic is a refinement of the hypothetico-deductive method of Theoretical Explanation, where we understand the latter as being in contrast with Genetic Explanation or Historical Explanation for instance.[38] We can see this by considering two comments on this form of theory by Wilfred Sellars:

> theories about observable things ... explain empirical laws by explaining why observable things obey to the extent that they do, these empirical laws.[39]

This is done by;-

> explain(ing) the behaviour of objects of a certain domain by 'identifying' these objects with systems of objects of another domain, and deriving the laws governing the objects of the first domain from the fundamental laws governing the objects of the second domain.[40]

Compare this with Pylyshn's comments:

> As ... realists we propose as the next step exactly what solid-state physicists do when they find that postulating certain unobservables provides a coherent account of a set of (observed) phenomena: we conclude that the (symbolic) codes are 'psychologically real', (and) that the brain is the kind of system that processes such codes ...[41]

---

[37] *ibid*, p.131-2.
[38] see Nagel (1961).
[39] Sellars (1963), p. 121.
[40] Sellars (1967), p.321..
[41] Pylyshn (1984), p.40.

In simple terms, 'heat is molecular kinetic energy', - molar properties are explained in terms of the properties of the unobserved microscopic particles of which they are composed. In the case of Pylyshn's model, symbolic processes, which are computational, instantiate the material basis for semantic level regularities, and thus explain the latter by allowing them to be derived from the principles of the former. In a sense semantic level regularities are what we *call* what are really the manifestations of an underlying symbolic level. Similarly, the symbol level regularities are explained and instantiated by the biological or physiological level. Thus the hierarchy proceeds 'downwards' something like this:

> Firstly:
>
> being the same thought entails having the same semantic content (that is, identical thoughts have identical semantic contents).[42]
>
> Then:
>
> to be in a certain representational state is to have a certain symbolic expression in some part of memory. That expression *encodes* the semantic interpretation.[43]
>
> or:
>
> symbolic codes ... reflect all the semantic distinctions necessary to make the behaviour correspond to the regularities that are stated in semantic terms.[44]

This statement, we will see, encapsulates the essence of a view of cognition which I shall argue is false.

Then we have at the physical level:

---

[42] *ibid*, p.43.
[43] *ibid*, p.29.
[44] *ibid*, p.39.

> the primitive functions or fixed symbol manipulation operations of the cognitive system.[45]

This is the 'Functional Architecture', the basic resources given in any particular programming language, which reflect all the symbolic distinctions necessary to make the system's behaviour correspond to the regularities which are stateable in symbolic/computational terms:[46]

> Mental architecture can be viewed as consisting of just those functions or basic operations of mental processing that are themselves not given a (symbolic) process explanation.

These are:

> instantiated in the biological medium.[47]

In short:

> Just as physical level principles provide the causal means whereby symbol level principles ... can be made to work, so symbol level principles provide the functional mechanisms by which representations are encoded and semantic level principles realised. The three levels are instantiated in an instantiation hierarchy, with each level instantiating the one above.[48]

STRONG EQUIVALENCE AND FUNCTIONAL ARCHITECTURE.

At the beginning of his 1978 paper 'Computational Models and Empirical Constraints' Pylyshn asks rhetorically:

> How close a connection can we expect between computational ideas and psychological theory ? Is the relation between the two to remain at the level of exchange of

---

[45] Pylyshn (1980), p.126.
[46] cf e.g. Pylyshn (1984), p.259-62.
[47] Pylyshn (1980), p.126.
[48] Pylyshn (1984), p.132.

> concepts and metaphorical focusing of attention or can it be more intimate ? In particular, can a program be a psychological theory ?[49]

Pylyshn's answer is 'yes': like Physics, computational psychology does not claim that its subject matter merely behaves *as-if* it follows its laws.

There is an ambiguity in the notion of computer simulation which reflects the relativity of theory to levels of description. For instance:

> When we simulate, say, the motion of planets, the only empirical claim we make is that the co-ordinate values listed on the print out correspond to the ones that will actually be observed ... Which alogarithm is used to compute these values is irrelevant to the veridicality of the simulation. In other words, for this purpose we do not distinguish among alogarithms that compute the same input-output function.[50]

A relationship of this kind between computational model and domain modelled is called 'weak equivalence'.[51] In Psychology Pylyshn wants to go beyond this and establish that not only may a computer program qua cognitive theory produce the same outputs from the same inputs in the domain so described, it may produce them in the same way. In computation this is to say that the same alogarithm is used by the brain as is presented in a true cognitive theory of that function. (As a first approximation, an 'alogarithm' is a completely specified procedure, or a mechanical 'recipe' for performing a particular task). This is the privileged vocabulary hypothesis again; cognition is alogarithmic and 'strong equivalence' is alogarithmic equivalence,[52] the imputation being that if a computer simulation is strongly equivalent to a cognitive domain then that simulation has the status of a cognitive theory which is also true:

> two programs can be thought of as strongly equivalent or as different realisations of the same alogarithm or the same

---

[49] Pylyshn (1978), p.94.
[50] Pylyshn (1980), p.119-20.
[51] cf *ibid*, p.117.
[52] cf Pylyshn (1984), p.90-1.

> cognitive process ... if they can be represented by the same program in some theoretically specified virtual machine."[53]

The appropriate level of comparison is somewhere between weak equivalence and complete identity because, of course, brains are made from quite different material from computers. Thus whilst the most basic operations of the brain will differ, it is implicit in the combination of the privileged vocabulary hypothesis (the theory) and the criterion of strong equivalence (the degree of equivalence between theory and subject matter) that the alogarithmic level of comparison is supposed to specify cognitive process down to the level of the most elementary functions that make a psychological difference. Accordingly, the privileged vocabulary hypothesis serves as an implicit definition of what it is to count as a cognitive phenomenon.

The alogarithmic level corresponds to what we have been calling the 'symbol level'. Computation is symbolic and computations are described in terms of alogarithms. Thus, as we have seen, ontologically speaking these alogarithms are instantiated by the non-symbolic hardware of the brain, the functional architecture. [54]

To the extent that we have so far elaborated these concepts it would seem reasonable to suppose that the functional architecture determines the range of possible mental alogarithms, and since the functional architecture is instantiated in the biological medium it would seem that a piecemeal bottom-up description of the brain - from neurons to groups of neurons to functional structures - must carve nature at the same joints as would a top-down account where beliefs and desires are explained by symbolic computations and these are in turn accounted for in terms of the biological architecture. Do we not mean by 'functional architecture' types of structures individuated by their physiological or mechanical kinds ?

Pylyshn cannot agree to this because if he allows that a bottom-upanatomical taxonomy of the domain makes exactly the same distinctions as a taxonomy beginning with the intentional explananda and working 'downwards' he is in danger of granting the physiological reductionist exactly what he wants, and so endangering the privileged vocabulary hypothesis. As the precise relationship between physiology and

---

[53] *op. cit*, p.91.

[54] 'Mental alogarithms are viewed as being executed by (the) functional architecture' ((1980), p.123.), the latter being the 'primitive functions that are themselves not explainable in terms of symbol manipulation processes' ((1984), p.30.).

symbolic states is of some importance to my argument we will labour this point.

Materialism.

The fundamental attraction of the materialist idea that "mental processes are nothing but a certain sort of physical process in the brain"[55] is almost to basic to state. A full analysis would perhaps begin by assimilating Truth to Rational Consensus, and then go on to characterise Science as more 'rationally consensual' than its competitors,- common sense, divine revelation, phenomenological scrutiny etc.[56] This yields the idea that the true, or most likely to be true, picture of Man and the World is the scientific picture, and true realism is scientific realism, whereby:

> in the dimension of explaining and describing the world, Science is the measure of all things, of what is that it is, and of what is not that it is not.[57]

This quote is revealing insofar as it implies that Science's being the measure of all things is equivalent to its arbitrating on questions of existence. The measurement is the measurement of what is and what is not. This makes substance the basic concept and ties it up with truth and knowledge. We can trace this notion back to Aristotle:

> of all these senses which 'being' may have the primary sense is clearly 'what a thing is' for this denotes substance, whereas nothing else is considered to exist unless by virtue of its being a quantity, quality, affection or other determination of substance.[58]

And this is why materialists seek to account for mental events in terms of a level of explanation which refers to what is clearly substantial like Physics or Physiology. Note also the difference between this attitude and that of Pylyshn quoted earlier, that the task of Psychology is to capture

---

[55] Armstrong (1980), p.34.
[56] see Armstrong, ch.3.
[57] Sellars (1963), p.173. For Scientific Realism see Smart (1963), Churchland (1979).
[58] Metaphysics, Book Z, I; Warrington (1967) p. 167.

generalisations (p25). Presumably Pylyshn would want to say that a theory which correctly 'captures the generalisations' is true, therefore we have two prima facie different attitudes towards truth, and goals of theory writing; to capture the appropriate generalisations, and to characterise what sorts of things the inhabitants of a given domain are.

FUNCTIONAL AND SEMANTIC LEVELS.

The denial of the reducibility of psychological generalisations to physics or physiology, and the necessity of an autonomous level of psychological explanation, is argued by Pylyshn in two ways:

1) It is stated that physiological or biological accounts would fail to deal with the problem under the intended interpretation; that is, they will fail to address the why, what and how questions posed by such cases as the pedestrian example in the same vocabulary as these questions. [58]

This is not a decisive objection because it implies that no framework of regularities can be explained except in its own terms. If this were true it would ensure that the properties of everyday objects such as tables, cars and ice cubes can have no explanation in terms of microscopic particles. In fact Theoretical Explanation is usually taken to refer to just this sort of explanation, where the regularities in one domain or level of description are accounted for in terms of the properties of another domain. This is how Hempel describes theoretical explanation:

> ... a theory construes (the phenomena to be explained) as manifestations of entities and processes that lie behind or beneath them, as it were. ... by means of which the theory then explains the empirical uniformities that have been previously discovered.[59]

So it is clear that if all phenomena can only be explained in terms of generalisations cast at the same level, then we cannot explain why I cannot pass my hand through my desk by referring to its microphysical structure. Nor can we explain why a burning object gains weight as it oxidises. As this is absurd it follows that it is quite appropriate to explain

[59] Hempel (1966), p.70.

molar psychological phenomena in terms of the qualities of a different category of objects.

In fact if the argument for the autonomy of person level explanation is taken to its logical conclusion it removes any need for the postulation of inner processes at all, even if they are computational. Wittgenstein has shown that it is quite possible to talk about beliefs, for instance, without having to suppose that there is any process or object within the confines of the person that *is* a particular belief.[60]

2) Pylyshn's main argument against reducibility is the need to capture generalisations ,"that is, that there be valid generalisations at one level that are not expressible at a lower level".[61] This leads him to postulate two distinct levels above the physical level, a symbol processing or functional level and a semantic or intentional level .[62]In the first case Pylyshn wants to say that there are generalisations capturable at the functional level which are not capturable at the neurophysiological level because:

> there is multiple realisation of functional states as neurophysiological states (thus) functional generalisations cannot, in general, be captured in a finite neurophysiological description. [63]

Secondly, we cannot rest here with a functional psychology which identifies mental states by their function within a closed system (the brain) because we must refer to the representational content of these states, and generalisations couched in terms of content will not necessarily coincide with even functional level generalisations. [64] These points are made via a comparison of content governed cogitation with the non-contentful but systematic operations of a wristwatch:

> Suppose I wish to explain what I am doing at this moment as I sit here in front of my computer terminal, alternatively

---

[60] see, for instance, his (1953), section 578.
[61] cf (1984), p.35.
[62] *ibid*, p.24.
[63] *ibid*, p.32-3.
[64] Pylyshn equivocates between 'symbol level' and 'functional level'. Strictly speaking - 'generalisations expressible in terms of functional properties of the functional architecture are referred to as "symbol level generalisations"'. (1984), p.32.

> typing these sentences and thinking about other things ... I think about the Santa Cruz mountains behind me and wonder why I didn't go for a walk after lunch instead of confining myself to my study. Clearly, my current behaviour (including writing this paragraph) is caused by my current thoughts and goals. They include the goal of having a completed chapter and the thought that there are walking trails in the hills behind me. If there is any sense in which this behaviour is caused by the non-existent completed chapter or by the hills, it is an entirely obscure sense of causation. Plainly what is going on is, my behaviour is being caused by certain states of my brain. Yet, - and this is the crux of the problem - the only way to explain why those states caused me to type the specific sentences about walking, writing, the mountains, and so on is to say that these states themselves are in some way related to the things referred to (writing, walking, mountains).[65]

The relationship between brain states and mountains will, for familiar reasons, be quite anomalous. Intuitively, it seems highly unlikely that everyone who glances at the Santa Cruz mountains will immediately go into a specific brain state and write at the earliest possible opportunity exactly the above words. It is only likely to be a rule that upon glancing at such and such mountains one goes into brain state X and makes such and such marks upon paper ceteris paribus, where the latter involves an almost infinite statement of initial conditions. Moreover, there is nothing intrinsic to a brain state that makes it about mountains, as opposed to hamburgers, boolean algebra or anything else. So in order to be able to state the rules, and in order to have any regularities to state them of, the brain states in question must be identified by, and with, contentful representational codes, allowing higher order principles such as rationality to come into play:

> the brain states must be treated as embodying representations or codes *for* such things as walking and hills. In this way the behaviour can be seen as rationally connected to the

---

[65] *ibid*, p.27.

> representational content of these codes by certain rules (logical, heuristic, associative etc.).[66]

But Pylyshn's further point here is that generalisations couched in terms of representational content refer to things extrinsic to the brain system and this taxonomy captures different generalizations from a description of the system in terms of the operation of its parts relative only to the rest of the system. [67]

The operations of a wrist watch, for instance, might usefully be described at the functional level. In particular, we can interpret the functions of the various cogs and levers in terms of the concept of time. Yet in the case of watches nothing is gained by this interpretation because the extrinsic/representational description will be type equivalent to a physical description. '12.05' will always coincide with just one arrangement of levers and cogs. For each time of day there will be just one physical description and for each physical description there will be just one (or two, am and pm) time of day, hence:

> the semantic interpretation is gratuitous, since the set of movements corresponding to the 'interpreted' behaviour is co-extensive with, or type equivalent to, the set of movements corresponding to the physical description of the behaviour.[68]

As watches don't have any other sort of behaviour apart from time telling, representational and functional descriptions of watches are co-extensive, hence by the principle of parsimony the representational account is otiose.

According to Pylyshn though, representational and functional taxonomies of brains and computers will only contingently coincide, i.e. they will often differ and thus capture different sets of generalisations. He has already stated that differences in content must always coincide with symbol level distinctions (p 40 above), and hence functional distinctions underlying these (and biological distinctions under these), so this contention rests on the assertion that:

---

66 *ibid.*

67 "generalizations can be captured by referring to the content of representational states that differ from those captured by referring to functional mechanisms." *ibid*, p.32.

68 *ibid*, p.28.

a) a given semantic/content level description may coincide with more than one functional level description, so that:

> There may be synonymous expressions - sets of codes with the same semantic content. Such codes might be functionally, but not semantically, distinguishable.

b) merely ascribing representational content does not suffice as an explanation, for the interactions of the codes so ascribed are left mysterious:

> Merely possessing a certain symbolic expression that encodes semantic content is, by itself, insufficient to produce behaviour ... What is needed is a set of mechanisms to make the system run or, ... 'interpret' the symbols. It is because of these mechanisms that the symbolic expressions do not exhaust what we have informally been referring to as the 'functional state' of the system.[69]

These mechanisms are what Pylyshn refers to as the 'functional architecture'. They include the basic biological operations for storing, retrieving, sorting and comparing symbols, as well as 'control structure' which selects which rules to apply at a given moment.

The implication of this last distinction is that the semantically interpreted symbols and thus the codes which they comprise are somehow discrete and autonomous, insofar as symbolic expressions seem to be able to maintain their identity despite different operations being performed upon them. It is as if each symbol were printed on a card, and thinking consisted of arranging them according to certain rules. So with the proviso that;-

> the mechanisms themselves are not sensitive to the content of incoming information, since, by hypothesis, semantic content is precisely what is encoded in terms of the symbolic codes,[70]

---

[69] both *ibid.*,p.30.
[70] *ibid*, p.31.

-he is free to suggest that the architecture may vary in ways unrelated to content whilst the representations remain unchanged. For instance, two persons in the same mental state content wise may nevertheless differ functionally if they have slightly different memory capacities, or retrieval mechanisms of different efficiency, resulting in different response times on certain tasks and different patterns of error and competency. Consequently:

> generalisations can be captured by referring to the content of representational states that differ from those captureed by referring to functional mechanisms.[71]

The final inference is that the autonomous psychological level of description corresponds to the subject matter of the independently developed theory of alogarithmic computation which:

a) charts a notion of process which is independent of physical instantiation, and,
b) engenders all of the level distinctions within Psychology which Pylyshn wants to make. To complete an earlier quote:

> (computer) behaviour is caused by the physically instantiated properties of classes of substates that correspond to symbolic codes. These codes ... (or symbols (which) are equivalence classes of physical properties) ... reflect all the semantic distinctions necessary to make the behaviour correspond to the regularities that are stateable in semantic terms.[72]

Thus brain and computational model are identified in the manner of theoretical explanation, viz.,

> ... the brain is the kind of system that processes such codes ...[73]

---

[71] *op. cit.*, p.32.
[72] *ibid*, p.39.
[73] *ibid*, p.40.

THE IMPLICATIONS OF PYLYSHN'S THEORY.

Having stated that a semantic level regularity may correspond to more than one functional level regularity, there is a sense in which the logical next step would be to argue that a given functional level regularity may take part in more than one semantic regularity, thus ensuring that the semantic and functional domains are doubly cross classified.However, Pylyshn disavows any suggestion that there might be participation by a given regularity or generalisation at level n in more than one regularity or generalisation at level n+1:

> Effects can penetrate upward through levels, since each level is supervenient on levels below; that is, there can be no difference at level n unless there is some difference at level n-1, even though the converse is not true (because of the multiple - instantiation property of ascending level; supervenience of psychological states on biological states entails that there cannot be two different thoughts unless there are *some* biological differences between the two underlying brain events).[74]

This reflects a consistent and important theme in Pylyshn's theory, that:

> differences in content always result in functional differences.
> [75]

or more to the point, as the theory aims to explain externally applied differences in content in terms of functional differences rather than vice versa (beliefs in terms of brain functions as opposed to brain functions in terms of beliefs):

> differences in content always and only *stem from* functional (symbol level) differences.

---

[74] *ibid*, p.38, his emphasis.
[75] *ibid*, p,29.

There is a sense in which this position could be irenic; a given neurological structure may coincide with a subprocedure that is called by two or more different symbolic procedures on different occasions. For instance, it is quite likely that speech comprehension and speech production have procedures in common. Thus two different 'thoughts' may coincide in part at the biological level. However, by his emphasis of the word 'some' above the implication is that Pylyshn takes this sort of case into account. He is thus not saying that different psychological states may not have some biological component in common, only that a given series of biological events cannot, in its entirety, be taken as identical to or co-exstensive with more than one 'thought', symbol or semantic level regularity. Psychology is determined by, but not reducible to, neurology. This bears upon Hilary Putnam's distinction between 'narrow' and 'wide' psychological states.

A narrow psychological state " ... presupposes the existence of (no) individual other than the subject to whom the state is ascribed"[76], all other states are "psychological states in the wide sense". Thus 'x is jealous of y' entails that y exists and is a psychological state in the wide sense. By these standards theories of mind which confine themselves to narrow psychological states are said by Putnam to make the assumption of 'Methodological Solipsism'. Insofar as Pylyshn is arguing that there can be no psychological distinctions without biological distinctions he is thus implying that narrow psychology is a complete psychology, i.e. that the organisms contribution - what is materially in his/her head - exhausts explanation in Psychology. But whilst this may be an *aim* of Pylyshn's, insofar as he also wants to say that psychological states possess 'of-themselves' a symbolic content he is in error. His 'symbol level' is narrow qua alogarithm, but wide qua symbol. Beliefs require mention of the propositions believed and the goals sought, which in turn often require mention of individuals other than the believing/desiring subject, -'John believes that the Eiffel tower is in France' etc. So to be precise, Pylyshn's attitude is that wide psychology is explainable by narrow psychology, or narrow-psychological facts, insofar as semantic level generalisations are explained by symbol level regularities, which in turn are equivalence classes of biological or neurological properties, which reflect all necessary semantic distinctions.

---

[76] Putnam (1975), p.220.

This contention is false. As I will show below, there may be semantic level distinctions which are reflected by no distinction in narrow psychological state. In other words there may be two different 'thoughts' with no biological differences between the underlying brain events. This is because the semantics of thought are determined in part by the state of the external world. To paraphrase Putnam, 'meaning is not (entirely) in the head'. Wide psychological states are underdetermined by narrow psychological states. This fact is significant because it is one of Pylyshn's central aims, and one of his central claims in opposing the analog school of imagery, that functional architecture can and should be determined by a chain of theoretical reasoning which begins with wide psychological, ascriptions. Witness to this is his contention that because response times vary in certain ways with informational factors in certain imagery experiments, because the tasks are 'cognitively penetrable', the functional architecture in question cannot be that of an analog data structure.

COGNITIVE PENETRABILITY AS A CLOSURE PRINCIPLE.

As we have seen, the assumption that psychological processes are alogarithmic amounts to the assumption that all genuinely psychological variability is alogarithmic,thus the privileged vocabulary claim amounts to an implicit closure principle. In similar fashion the hypothesis that heat is molecular kinetic energy implies that any genuine change in temperature is, or involves, a change in molecular kinetic energy.

> ... the privileged vocabulary claim ... asserts that cognitive phenomena can be accounted for solely by appealing to the symbolic representations (i.e. the alogarithm and its associated data structures). Thus, any differences among such phenomena arise solely from the structure of these symbol systems - from the way the component parts are put together. Thus, no distinctions among phenomena that we would class as cognitive distinctions can be due to such things as differences in the way the primitive operations that constitute the alogarithm themselves function.[77]

---

[77] Pylyshn (1980), p.126.

In accordance with his strong equivalence claim, the hypothesis in question is not that cognitive processes are *like* alogarithmic processes, but that cognitive processes *are* alogarithmic processes. Consequently the claim that is being made is one of contingent identity - contingent because it is possible for the theory to be false; i.e. cognition may turn out to be only explainable as a neurophysiological function of the brain. Given that there are certain phenomena that are extra or pre-theoretically identifiable as 'cognitive' (mental arithmetic for instance), the general form of Pylyshn's argument for cognitive penetrability is something like this:

> If what is pre-theoretically judged to be an alteration in cognitive initial conditions - such as a change in the information the subject possesses - produces an alteration in output (with respect to previous outputs) then the function in question is not attributable to the fixed biological structure of the brain.

The argument is at its most concise and complete in his 1981 paper. Firstly he defines analogue process:

> we would count (a) process as analogue if its going through particular intermediate stages were a necessary consequence of intrinsic properties of the mechanism or medium, ...[78]

Whereas the tacit knowledge account of imagery phenomena:

> claims that how the representation will behave is a function of what the person knows about the actual behaviour of the things represented, rather than of properties of the medium in which it is represented.[79]

i.e. the subject knows that certain distance-time effects obtain in the world and so 'makes it the case' using some form of symbolic computation, that these relations will hold when he/she imagines the events. In general, as we have seen, a large proportion of behaviour must be explained in terms of beliefs and goals, - rules and representations, -

---

[78] Pylyshn (1981), p.20.
[79] *ibid.*

because the regularities of interest are physically and behaviourally anomalous. Pylyshn continues:

> A corollary of this explanatory claim is that if a certain behaviour pattern (or input-output function) can be altered in a way that is rationally connected with the meaning of certain inputs (i.e. what they refer to, as opposed to their physical properties alone), then the explanation of that function must appeal to operations upon symbolic representations such as beliefs or goal: It must, in other words, contain rule-governed cognitive or computational processes. A function that is alterable in this particular way is said to be cognitively penetrable.[80]

Again:

> A process that is sensitive to the logical content of beliefs must itself contain at least some inferential (or other content dependant) rule-governed process.[81]

On the other hand if the analogue approach were true:

> ... manipulation of such things as the form of the task and the instructions should not have a corresponding, rationally explainable effect (provided, of course, that imagery was still being used). Otherwise we would have to say that the medium changes its properties to correspond to what subjects believe about the world, in which case appealing to the existence of an analogue medium would serve no function.[82]

In other words, if the analogue spatial medium proved to be cognitively penetrable it would ipso facto cease to be an analogue medium, but something more akin to a 'symbolic representation' in Pylyshn's sense, leaving the analogue theorist with the choice of either a), abandoning his model completely or b), admitting symbolic components

---

[80] *ibid*, p.21.
[81] *ibid*.
[82] *ibid*, p.23.

to the model so that it ceased to be analogue in anything but name - for Pylyshn allows that symbolic processes may have analogue *components* ,[83] hence any process which is symbolic in part is ipso facto not *analogue.* That is, cognitive penetrability is a sufficient condition for a process to be symbolic. A composite cognitive function is a *symbolic process with an analogue component,,* not an *analogue process with a symbolic component.*

One point at which Pylyshn's argument is quite secure is his definition of an analogue medium. Once Kosslyn's point is taken, that the analogue claim refers only to a subset of the cognitive processes involved - the visual buffer- he shows no signs of disagreeing with Pylyshn's characterisation of the medium as having intrinsic properties, and accepts that this part of his model should not be cognitively penetrable.[84] Whilst there is some uncertainty as to the intelligibility of a medium that is only functionally spatial, -i.e. possesses no literal spatial extent,-[85] yet is not propositional, Pylyshn's use of the word 'intrinsic' seems to confer neither more nor less residual propositionality to the medium than Kosslyn's description of it as 'innate'.

Pylyshn's general point against the purported significance of the imagery literature taken as a whole takes as its point of departure a distinction between two types of explanation, a distinction motivated by his conception of cognition as involving three discrete levels.

The first of these consists in the appeal to "symbolically encoded facts about the world and to rules for transforming representations and drawing inferences." That is, explanation which involves just the resources of the 'symbol' and 'semantic' levels, whose biological instantiation is of no concern to Psychology. The second type of explanation adverts to "intrinsic lawful relations among properties of a particular physical instantiation of a process,"[86] or what Pylyshn would call the Functional Architecture, which is the type of explanation relied upon in Kosslyn's (and others) accounts of mental scanning and image rotation effects.A distinction is then made upon this basis between two quite different tasks which, a priori, the subjects of imagery tasks could be seen as performing:

---

[83] *ibid,* p.21.
[84] Kosslyn (1981), p.56.
[85] *ibid,* p.50.
[86] Both Pylyshn (1981), p.19.

1) Using a mental image, focussing attention on a certain object in the image and deciding as quickly as possible whether a second named object is present in that image, or,

2)'Imagining' yourself in a certain real situation in which you are viewing a certain scene and are focussing directly on a particular object in that scene. Then 'imagining' that you are looking for (scanning towards, glancing up at, seeing a speck move across the scene towards) a second named object in the scene. And then when you succeed in 'imagining' yourself finding (and seeing) the object (or when you see the speck arrive at the object), performing a specified task.[87]

Now inter alia, the performance of task two will require the accurate recreation of all the requisite response latencies, i.e. the image scanning and rotation effects detailed in chapter one, as part of the 'task demands' of the situation.[88] But this does not necessarily require that the subject actually use images qua analogue mediums. Subjects are able to successfully perform the required tasks, in Pylyshn's view, because they possess *tacit knowledge* of the way things actually happen in the world:

> e.g. ... the subjects would know implicitly that, for instance, it takes a moving object longer to move through a greater distance, that it takes longer to shift one's attention through greater distances (both transversly and in depth).[89]

This, in turn, is possible because we know independently that human subjects have the ability to generate time intervals corresponding to known events, and that this mechanism need not have anything to do with any purported imagery medium (see Fraisse (1963), Chapters 5 and 7).

Thus the numerous studies which purport to demonstrate the existence of an inner picture-like medium come out as they do because subjects interpret their task as being the second one detailed above, and because they possess an independent ability to generate all the required response latencies. Consequently it is alleged that these studies do not

---

[87] Paraphrase of Pylyshn's tasks 2a and 2b, p.235, (1984).
[88] cf Newell and Simon (1972).
[89] Pylyshn (1984), p.236.

provide demonstrative proof that anything like a mind's eye or imagery medium is involved (cf Pylyshn (1984), p.235-7, Kosslyn (1981), p.61).

Whilst this is a plausible counter-explanation of the phenomena in question, it is nevertheless just as non-demonstrative as the analog view. It is possible that subjects are performing task 2 instead of task 1, but how can we be *certain* that they are performing task 2 ? It is difficult to imagine what would count as proof either way.

TACIT KNOWLEDGE.

There are, I believe, serious problems with the tacit knowledge view as long as it is thought of as being an autonomous level of explanation. Insofar as it might be said that a sufficient explanation of a psychological function can be given just by citing the semantics of its domain, then this account is mistaken. The problem comes down to the fact that whatever we can cite as someone's tacit or explicit knowledge of something is something which we already know. Consequently it is not so much the explanation as the thing to be explained.

Suppose we explained the fact that a very large proportion of the population will respond '5' to a request to add '3' and '2' by saying that they *know* that '2' and '3' are '5'. But if the semantics of mathematics were different, which is just to say that if normal (strictly speaking) human practice had evolved differently over the ages, then then the normal response to this question could be a number which sounds like '6' or '23'. Consequently the explanation that people respond with '5' because they *know* that 2 plus 3 makes 5 is not counterfactually supporting , an d as Pylyshn himself maintains, counterfactual support is a necessary condition for explanation as opposed to mere description (see p.6, (1984)). That what people 'know' is to a large extent relative to culture also shows that knowledge, whether tacit or otherwise, is not an individual capacity. W h

at a person can be said to know is relative to socially determined truths.

TASK 1 OR TASK 2 ?

Cognitive penetrability aside, there are no uneqivocal criteria for determining which of these two tasks a given input-output function may instantiate.

Even if there is no explicit coaching to 'imagine yourself in a certain real situation' any element of visual pictorial stimuli can be taken as an invitation to do similar. Indeed, the only relatively conclusive evidence that tacit knowledge is not behind a given experimental result is the demonstration of 'tacit ignorance', and the lack of any 'image scanning' correlations between time and distance.

Kosslyn replies to the assertion that tacit knowledge is the appropriate account for putative imagery phenomena by insisting that a), subjects are not in any way exhorted to use imagery; b), imagery effects emerge just for those properties which are independently assessed as requiring imagery, and c), imagery and perception share certain counter-intuitive properties quite at odds with subjects expectations (cf Kosslyn (1981), p.61-3). But these points can all be countered:

a) Pylyshn makes it clear that what is at issue is not the extraneous contamination of the experimental task by the experimenter's instructions to the subjects, but what we have come to know as the Task Demands, where "subjects (solve) a task as *they* interpret it".[90] Nothing needs to be said to the subjects at all, they simply bring to bear the capacities they feel appropriate to the task. One does not have to be told or instructed to use one's chess knowledge when given a chess problem.

b) In Kosslyn, Jolicoeur and Fleigel,[91] to take an example, subjects were asked to imagine an object and 'mentally stare' at one end of it. They were then asked if the object possessed a certain property. Verification times, it turned out, were proportional to the distance from focus point to the property on the image (the bee's head for instance) only for those properties that a separate group of subjects had rated as requiring imagery to verify. But if the use of tacit knowledge is regarded as self initiated in the face of appropriate stimuli, then the only significant factors are the nature of the stimuli and the kind of questions asked or tasks given. Instructions, or the lack of them, are of no consequence. It is a simple product of the fact that the tacit knowledge in question is tacit knowledge of certain geometrical principles and time and distance relationships that the familiar response latencies are elicited only by questions about properties that fall within this domain. On any account of cognition a

---

[90] *op. cit.,.*

[91] Described in Kosslyn *op cit*, p.61.

question like 'Does the honeybee have a dark head ?' will access a different kind of knowledge from 'Is the honeybee an animal ?' just *because* they are different kinds of knowledge. The response latencies are data-driven, hence it is the spatial extension of the bee itself - as opposed to representational media - which cause certain types of response delays in most people, and are thought of as requiring imagery by most people, including *all* the experimental subjects.

c) The third strand of Kosslyn's objection rests upon the assertion that the sorts of results obtained when subjects 'consult their images' differ from their previously avowed beliefs and expectations. Thus in Finke and Kosslyn (1980) it is established that the point at which two dots merge, when subjects are asked to imagine them receding into the distance, is consistently misjudged by a group of controls who are set the same task but asked not to use imagery. Also the width of the imagery field was found to be 1.83 times larger than the estimate given by the control group. In another experiment summarised in his (1981) paper Kosslyn reports that when subjects are asked to imagine a grating of black and white stripes receding into the distance, vertical stripes will blur at a greater distance than oblique ones, an effect of which none of the subjects knew beforehand. How can the tacit knowledge account get around the fact that, in general, people's explicit beliefs and prior expectations will not always match the results achieved by 'image inspection'?

The simple reply is that 'tacit knowledge' is *tacit.*, hence implicit. This is how Fodor defines tacit knowledge in an early paper:

> if an agent regularly employs rules in the integration of behaviour, then if the agent is unable to report these rules, then it is necessarily true that the agent has *tacit* knowledge of them.[92]

Tacit knowledge is by hypothesis not explicit, and so not accessible to consciousness or verbalisation. A good example of this are the rules of grammar; the linguistic competency of most people far exceeds their ability to state the rules which govern that competency. By the same token it is not in the least surprising that subjects verbally expressed prior beliefs and expectations do not always match up with the results they or other subjects

---

[92] Fodor (1968a), p.636, his emphasis.

obtain when the tasks in question are actually performed. As far as reaction time goes this result is underlined by Fraisse's report that, in general, people are far more accurate when they reproduce a given duration (i.e. upon being asked to write for one minute) than they are at verbally giving the probable duration of an event (i.e. that 'writing this piece took one minute).[93] Thus when Kosslyn argues that the regularities which appear in the experimental literature are often not known to or anticipated by the subjects, Pylyshn is able to reply that subjects are recalling and inferring from various past perceptions of actual objects, and that;

> It is true of recall in every domain that relevant facts frequently can't be accessed without going through some particular sequence or access cues.[94]

This is clearly related to the above report that duration reproduction is far more accurate than duration estimation.

In general, there would appear to be no limit to the ability of access limited tacit knowledge to account for non-intuitive effects in imagery experiments. This is underlined by Pylyshn's comment that tacit knowledge,-

> could obviously depend on *anything* the subject might know or believe concerning what usually happens in the corresponding perceptual situations.[95]

On the other hand Pylyshn's criticisms of Kosslyn's position are just as easily met. Two of the formers most convincing counter examples are described in his (1979) and (1981), (1984).

The first study involved two experiments very much in the tradition of Cooper and Shepard (1973) and Shepard and Metzler (1971). Subjects were required to rotate and match probes with a stimulus figure. 'True' probes, as opposed to distractors, varied in their goodness of fit in the gestalt sense (as measured by independent criteria) with the stimulus figure. The finding of significance was that the slope of the relationship

---

[93] cf Fraisse, *op cit*, p.210-13.
[94] Pylyshn (1984), p.246.
[95] ibid, p.34, his emphasis.

between the relative orientations of the figures to be matched, and reaction time, varied from probe to probe as a function of the gestalt 'goodness' of the embedding. On the face of it goodness of fit should effect only the final matching stage of the comparison. Thus whilst we could expect variations in the y-intercept from probe to probe, we would not expect the slope or 'rate of rotation' to be effected.

On the tacit knowledge view figural complexity and the difficulty of the post-rotation comparison task could constrain reaction time in any one of a number of ways, the theory having no particular prediction on this front. But in a model which relies upon an autonomous non-conceptual depictive medium, we would not expect any property of the image per se to effect the rate at which it is rotated.

> (this study) provides strong evidence that the process is not one in which a stage of holistic analogue rotation of the image is followed by an independent stage of comparison ... (these results) make it clear that if there is anything which might be called 'rotation' in this situation, the whole figure is not carried along rigidly. Rather, there must at least be some analysis of the original stimulus and some piecemeal 'rotate and compare' subprocesses.[96]

Whilst a piecemeal rotation model lacks the purity and intuitive appeal of the holistic 'image as a whole conception', this is of course exactly the tactic Kosslyn uses to reply to Pylyshn's criticism:

> ... perhaps the subjects did not encode the entire figure into a mental image but encoded only parts that they hoped would help in performing the task. If they guessed wrong, they fixated again on the figure and re-parsed it, encoding different parts into the image. In this case, when the test part corresponded to a 'bad' part of the figure (one that violated natural parsing procedures), subjects would have to encode the figure many times and encode it each time. Thus, the effects of angular disparity would be more pronounced for 'bad' parts (and the difference in slopes would reflect the

---

[96] Pylyshn (1979), p.27.

> number of times the figure was re-parsed to find the 'bad' part).[97]

The second study involved two experiments which followed closely the design of Kosslyn, Ball and Reisser (1978). Subjects were asked to memorise a map containing a number of visually distinct objects - a beach, a castle, a church and so on. They were then asked to image the map and to concentrate their attentions on one of the named locations, whilst keeping the rest of the map in view in their 'mind's eye's'. Initially subjects were asked to scan to a second named location on cue, by imagining a speck moving across the map from the first location to the second. This of course resulted in neat correlations between imaginal distance and reaction time. Subsequently, however, subjects were simply asked to give the compass bearing of the second location, in which case all correlations between distance and time disappeared. This did not rule out the possibility though that subjects used a symbolic representation since even in the first experiment the subjects must know the direction of the second location if they are able to scan towards it. Thus in the second condition subjects were required to first focus on the second place after they heard its name, and from this vantage point give the orientation of the first place. The instructions stressed the need to see both places before making the orientation judgement. The results yielded no significant correlation between reaction time and distance between places, underlying the earlier result, and implying that subjects performed the task by consulting a symbolic representation of the scene, rather than scanning across the image.[98]

Kosslyn's reply to this is to maintain that scanning need only be employed when the part of the object in question is depicted too far towards the periphery of the imagery field to be easily discerned:

> One ready account of Pylyshn's finding that subjects could judge relative orientation of imaged objects without scanning among them rests on the fact that one can 'see' more than a single location in an image at the same time. Image inspection is not like viewing an object through a small hole in a piece of cardboard that must be moved around to infer a

---

[97] Kosslyn (1981), p.58.
[98] cf Pylyshn (1984), p.243-4.

> general shape. Thus, subjects conceivably could have performed the Pylyshn task without having to scan.[99]

A second independent counter-explanation mentioned by Kosslyn is that it was found to be necessary to associate relative location information with any object or part of object imaged, because it was known that subjects could scan to locations off the imagery field. This being true of the objects in the Pylyshn experiment, it could well be that subjects performed the task without specific recourse to their images, so producing the results obtained. Kosslyn could also claim that subjects were using 'blink' transformations, a procedure allowed by the analogue model in which transformations are equally easy and fast over different distances.[100]

With a plethora of ready replies to counterexamples at hand it is not surprising that Pylyshn remarks "... it is hard to think of any result which could not be naturally accommodated (by the analogue model)".[101] Yet his tacit knowledge account is equally ubiquitous as Kosslyn notes;

> ... one can never be sure one has controlled for the effects of tacit knowledge in an experiment, in Pylyshn's view. In other words, whereas the demand characteristics account may be disprovable, Pylyshn's account, resting on implicit task demands, is sheltered from such a rude fate.[102]

As it stands the issue is to all intents and purposes an impasse. The one genuinely powerful argument though is the cognitive penetrability criterion. Although the analogue imagery model is still adhered to by some, strictly speaking the penetrability argument cannot be met.

THE STRENGTHS OF COGNITIVE PENETRABILITY.

1) This criterion is a straight forward extension of the core principles of the dominant general theory of cognition in modern Psychology, thinking as symbolic computation. One of the most central distinctions in this area, from which the cognitive penetrability criterion is derived, is the

---

[99] Kosslyn (1981), p.59.
[100] cf *ibid*, p.59-60.
[101] (1981), p.42.
[102] Kosslyn (1981), p.63.

distinction between hardware and software. If the symbol level is just that part of the system which varies in accordance with what we can roughly call the semantics or meanings of 'thoughts', and the functional architecture is just that part of the system which is fixed with respect to the symbol level, then it is difficult to avoid the conclusion that no putative psychological function which varies with the semantics or meaning of its inputs can be instantiated in the functional architecture. The analogue theory is not similarly motivated by a general theory of cognitive function. There is nothing like the analogue approach to concept learning for instance, whereas the computational approach has clear implications for this area.

2) In many ways the case for penetrability is understated. On a strict reading of the criterion it is inappropriate to even consider an analogue model for imagery, regardless of the evidence, for the model postulates architectural functions which vary in accordance with the meanings of inputs. In the literature the argument has proceeded as if the central aim was to establish that certain subprocesses within the model such as scanning or rotation were penetrable or impenetrable but this is already to give all the ground that counts to the analog position. If any part of the analog model is provisionally allowed to retain cognitively penetrable architectural processes the theoretical motivation for showing that other parts of the model are penetrable would be lost. If the penetrability principle is anything less than exclusive Kosslyn has already won his case, for he is concerned to show that only part of his model is cognitively impenetrable, and his theoretical background makes no caveats against a mixed model:

> ...the existence of non-analogue components in a given set of processes in no way bears on the truth or falsity of the claim that one component is an analogue spatial medium ...[103]

Pylyshn's theoretical position however demands that architectural structures play no active part in psychological theory, just as the chemical structure of genes plays no part in mendelian genetics. Hence the analogue model contravenes basic cognitive principles right from its conception. The only way to counter the force of this argument is to show that the cognitive penetrability criterion is theoretically ill founded, or is itself at

---

[103] Kosslyn *op. cit.*, p.57.

odds with certain empirical results. No debate as to whether tacit knowledge has to strain to account for certain counter-intuitive results in imagery or perception, or whether the apparent penetrability of a given process may be localised allowing adjacent functions to remain analogue, is to the point if the principle itself is unchallenged. Given that there is no question that imagery phenomena qua input-output function are cognitively penetrable, all discussion is pre-empted and the debate is resolved.

If on the other hand, as I will show, the cognitive penetrability criterion is seriously flawed, the debate is just as surely resolved in favour of the analogue camp, for by Pylyshn's own lights functional architecture is the preferred level of explanation for all cognitive phenomena as long as it can capture the relevant generalisations (cf p.38 above). So if Kosslyn's theory is a) preferred on these grounds, b) fully and specifically accounts for the empirical data and c) contravenes no general principles of psychological explanation, this view must prevail. If the penetrability principle is misguided the distinction between symbolic and architectural levels of explanation will break down, which must effect the integrity of any supposed domain of tacit, symbolic, knowledge. If we can say of imagery 'it isn't tacit knowledge it's an analog/architectural process' then it becomes difficult to see what could count as a criteria to the effect that the explanation of such and such a cognitive domain *must* and *can only* be explained in terms of symbolically instantiated tacit knowledge of 'the way things are in the world'. All 'symbolic' cognitive explanation becomes provisional and liable to replacement by more parsimonious biological accounts as soon as they can be developed.

A DIGRESSION ON THE FORCE AND SCOPE OF THE COGNITIVE PENETRABILITY CRITERION.

Before continuing we must be clear on the applicability of the criterion. Two queries present themselves:

1) Is not the mere intelligibility of Kosslyn's model a counter-example to the penetrability criterion, thus disproving it ?

In a sense yes, because it shows that in principle an erstwhile 'symbolic' process *may* be accounted for otherwise, thus showing that the symbolic

account is not necessary (in the logician's sense of 'necessary') as its proponents argue. But arguing for a certain conclusion by presupposing it clearly begs the question, and so is an inappropriate tactic in the present discussion.

2) The discussions in the literature point top the issue being as to whether certain cognitive functions qua 'processes in the head' are penetrable/impenetrable. Consequently the application of the label 'cognitively penetrable' to a domain on the basis of behavioural evidence always leaves it open that certain processes underlying the behaviour are not cognitively penetrable even when the range of behaviours in question are manifestly and unequivocally penetrable. Pylyshn even explicitly accepts this possibility when he writes:

> It should also be noted that being cognitively penetrable does not prevent a process from having analogue components: It simply says that it should not be explained solely in terms of analogues with no reference to tacit knowledge, inference, or computational processes.[104]

But given that Kosslyn allows such composite models (page 65) there seems to be no reason left to decry analogue spatial media.

Let us call the interpretation of cognitive penetrability which would allow us to judge a domain in general (such as 'imagery tasks') as penetrable on the basis of an observed co-variance between the semantics of the inputs and behavioural output, and thus having no significant behavioural part, the strong interpretation of cognitive penetrability, or 'SCP'. Let us call that interpretation of the criterion which allows that a broad co-variance between input-output behaviour and the semantics of stimuli may be consistent with some significant theory relevant part of the process being analogue, the weak interpretation of cognitive penetrability or 'WCP'. Kosslyn obviously cleaves to WCP, but Pylyshn may be read as supporting either. Despite comments such as the one above there are a number of reasons for taking SCP to be the received interpretation.

Firstly, the WCP interpretation makes the criterion toothless. If;-

---

[104] Pylyshn (1981), p.21.

> ... the point of the cognitive penetrability condition is to provide a purely functional methodology for deciding whether a putative property qualifies as belonging in the category of architecture or in the category of cognitive process,[105]

-then this aim cannot be achieved if every behavioural demonstration of penetrability is consistent with some significant part of the cognitive process being performed by an analogue or architectural component. The postulation of implausible but conceivable analogue components in cognitive processes would seem to be limited only by the imagination of the theorist. Thus we might suppose that reasoning and conceptual thought are achieved by the application of logical operations to sentences in English inscribed on some inner screen. Arithmetic could be performed by some sort of neural abacus. We feel sure that examples of this kind are nonsense but what is the difference in principle between these and sensible analogue using models such as Kosslyn's ? If the condition cannot preclude the existence of significant analogue components in models of cognitive process, then it is powerless against even the most ridiculous of analogue using theories.

Secondly, in Pylyshn's own terms his equivocation stems I think from his acknowledgement that there are conceivable cognitive models which are nevertheless lacking in explanatory adequacy. This principle emerges from his discussions of Anderson's indeterminacy of representation claim:

> (the) general point is that it is not possible for behavioural data to uniquely decide issues of internal representation. The reason is that one cannot just test questions about a representation in the abstract. One must perform tests of the representation in combination with certain assumptions about the processes that use the representation. That is, one must test a representation-process pair. One can show that given a set of assumptions about an image representation and a set of processes that operate on it, one can construct an equivalent set of assumptions about a propositional representation and its processes. Or one can be given a

---

[105] *op. cit.*, p.42.

> propositional theory and construct an equivalent imagery theory. (In general) Given any representation-process pair, it is possible to construct other pairs with different representations whose behaviour is equivalent to it. These pairs make up for differences in representation by assuming compensating differences in the processes.[106]

Anderson creates an existence proof for a cognitive model described in terms of processes E*, T*, and D*, by defining each of these processes in terms of another model described in terms of stimuli S, encoding functions E, internal representations I, internal representation transformation functions T and representation to response decoding functions D, which by hypothesis use different representations, yet map the same stimuli to the same response. Once a one to one mapping f is postulated which maps the encodings of stimuli in the target theory onto the encodings of stimuli in the mimicking theory - i.e. the internal representations of one theory onto the internal representations of the other - Anderson alleges that internal distinctions have been preserved such that the two theories assign corresponding representations to the same stimuli. With f in hand he also possesses the resources to define E*, T*, and D* in terms of the first model in such a fashion as to "prove" that the second model must map the same stimuli to the same responses as the first model. Thus:

$$E^*(S) = f(E(S));$$

$$\text{for each } T^*,\ T^* = f.T.f^{-1};$$

$$\text{and for each } D^*,\ D^* = D.f^{-1}.$$

Pylyshn's point though is that:

> ... even if some phenomena could be predicted equally well by either of two different representation-process pairs, one would not be licenced to conclude that the two systems were indistinguishable in general or ... that they would lead to equally explanatory theories.

---

[106] Anderson (1978), p.262-3.

He goes on to say:

> One of the things that makes a certain use of a model explanatory is that it appeals to general principles and mechanisms ... Unless the mechanisms posited in an account of experimental findings have some independent motivation other than to account for the data at hand, they can justifiably be accused of being ad hoc.[107]

Thus whilst Anderson's feat of mathematical gymnastics is quite clever, it is not succinct, recalling that by the standards of strong equivalence the computational complexity of a model is psychologically real if that model is true. The most explanatory model of a cognitive process is the least complex and most principled one that fits the data. Anderson's model is needlessly complex and unconstrained by any psychological principles. By the same token a maximally explanatory model cannot countenance arbitrary analogue components. A given analogue representation and process pair may be made formally equivalent to a second non-analogue representation and process pair by compensating for the rigidity of the analogue representation (part of the functional architecture) with additional flexibility of its accompanying process. But such a tactic would violate the original reasons for postulating a fixed functional architecture with respect to the symbol level in the first place, which is to provide a mechanism for the manipulation of token symbols. Thus once the conceptual geography of a domain is mapped out the architecture is arranged around (as it were) the symbolic interactions posited in explanation. Thus we cannot allow the architecture/symbol balance to be continuously variable in order to prove some obscure formal point. The most principled interpretation is that all input-output behaviours judged as symbolic or involving meaning are performed (only) by the symbol level.

Thirdly, and most significantly, to say that behaviour is representation governed in Pylyshn's theory is to apply to it the privileged vocabulary hypothesis, which states that 'cognitive phenomena can be accounted for solely by appealing to the symbolic representations (i.e. the alogarithm and its associated data structures)'(p.53 above). But the criteria for appealing to

---

[107] Both Pylyshn (1979) p.384.

a representational and thus alogarithmic account are precisely those for judging a process to be cognitively penetrable. We say that a behaviour pattern is cognitively penetrable if it may be altered in a way that is "rationally connected with the meaning of certain inputs" (p.55 above). But the regularities which we seek to capture with a symbolic computational theory, which cannot be captureed in behavioural or physiological terms are those of informed intelligent behaviour, acting upon beliefs, desires and deductive reasoning.

Pylyshn gives three criteria for behaviour to be explained by rules and representations:

> ... we would describe the behaviour as being governed by representations and rules if the relation between environmental events and subsequent behaviour, or the relations among functional states themselves, could be shown to be ... a) arbitrary with respect to natural laws, b) informationally plastic, or c) functionally transparent.[108]

Now, an informational effect upon a cognitive process or behaviour, - being an effect which is of course 'rationally connected with the meaning of the inputs' - will always be arbitrary with respect to natural laws. I can tell you to meet me at 7pm, or write you a note, or tap it out in morse on the table - and each of these in a different language. None of these will have sufficient physical commonalities to partake of the same natural law. Neither can there be any natural law relating what I say, when this is just described as a noise of a certain sort, with what you subsequently do, for that will depend upon your beliefs and expectations.

The second condition of informational plasticity is so similar to cognitive penetrability that it needs no comment:

> ... epistemic mediation ... is implicated whenever the relation between environmental events and behaviour can be radically, yet systematically, varied by a wide range of conditions that need have no more in common than that they provide certain information ...

The last criterion of functional transparency,-

---

[108] Pylyshn (1980), p.120.

> reflects the multiple availability of rules governing relations among representational states. Wherever quite different processes appear to use the same set of rules, we have a prima facie reason for believing that there is a single explicit representation of the rules, or at least a common shared subprocess, rather than independent identical multiple processes.[109]

This is on a par with the rational variability of cognitively penetrable processes, for each and every one of a set of processes which 'appear to use the same set of rules' is by hypothesis following or varying in accordance with some rule. We know from the context that the rules in question will be rules of symbolic computation, and thus that their variability will be, roughly speaking, 'semantic' or 'rational'. Consequently, the processes in question will vary in a way rationally or semantically connected with the meanings or semantics (etc.) of their inputs, whether these are sensory stimuli or other cognitive processes.

Thus the conditions for a cognitive process or behaviour pattern being cognitively penetrable are identical to the conditions for a behaviour pattern or cognitive process being governed by, and thus properly explained in terms of, rules and representations. But once this is granted we are governed by the privileged vocabulary hypothesis which maintains "that cognitive phenomena can be accounted for solely by appealing to the symbolic representations" (p.53 above). And explanations are by their very nature general (we don't say that 'most heat is molecular kinetic energy, but some of it isn't'), thus all cognitively penetrable behaviour patterns must be explained *solely* by appealing to symbolic representations, and this is SCP.

Thus the privileged vocabulary hypothesis is a closure principle, as is the cognitive penetrability condition; both engender a theory of cognition, but they also define what is to count as a cognitive process. This is not surprising because the principle of theoretical explanation is, roughly, to explain overt phenomena by hypothesis as to what the underlying processes 'really are' (cf p.39 above).

Perhaps one could try to hold on to WCP by arguing a degree of indeterminacy or unavoidable error in the application of either the

---

[109] Both (1980), p.121.

cognitive penetrability criterion or the conditions for representationality. It could be allowed that unavoidable indeterminacies in the application of the conditions allow cognitive penetrability to be adjudged of cognitive processes which are in point of fact not 100% alogarithmic/symbolic.

There are two ways of interpreting this charge: In the first case we could take it to engender a statement like 'conditions a-c above and the cognitive penetrability criteria are 100% acceptable, but they cannot be applied with 100% accuracy.' This is an epistemological point which bears upon the problem of inductive indeterminacy which is shared by all scientific theories. It is always possible for more than one theory to be constructed which is consistent with any finite set of data. As Popper has argued[110] there is a sense in which no amount of evidence can conclusively verify a theory. No matter how many black crows have been observed it is always possible that there are white crows that nobody has seen. Although it has never been observed, lead occasionally turns to gold. But this indeterminacy should not effect the claims of the theories themselves (not at least since the demise of Logical Positivism). We say, 'all crows are black, though we are only 98% sure of this', from which it does not follow that 98% of crows are black. Similarly, once the appropriate conditions are satisfied, there will be no non-symbolic content amongst the underlying cognitive processes. Although the conditions cannot be applied infallibly, the theoretical claim is that *if* correctly applied, the underlying processes are determinately as the theory says they are. As Pylyshn notes in this respect Psychology is in the same boat as the physical sciences, thus the indeterminacy involved in ascribing mental representations is (allegedly) no more than that involved in ascribing hypothetical entities in Physics.[111]

To conclude; it is clear that the interpretation of the Cognitive Penetrability condition which follows from a detailed examination of Pylyshn's conception of cognitive process is the strong interpretation, 'SCP'. From this it follows that *as long as the cognitive penetrability condition is valid,* any domain judged as being cognitively penetrable (epistemological considerations aside) can have no significant architectural and hence 'analogue' component. In other words there is no way around cognitive penetrability - this having been the predominant tactic of the analogue school. One cannot leave cognitive penetrability and

---

[110] cf Popper (1929).
[111] Pylyshn (1979), p.383.

the associated criteria for 'representationality' uncriticised whilst trying to say that there is, nevertheless, room for analogue processes. Kosslyn's theory does not have any general conception of cognitive explanation to wield against the exigencies of Pylyshn's system; in this situation the systematic theory must inevitably prevail. The argument looks like this:

What is cognitive is (identical with what is) alogarithmic.

What is cognitively penetrable is cognitive (by the identity of conditions a-c above with the criteria for cognitive penetrability).

Therefore, what is cognitively penetrable is identical with what is alogarithmic.

Therefore, if a process or behaviour is cognitively penetrable it can have no significant architectural or analogue component.

Therefore, either the analogue theory of imagery is false, or the cognitive penetrability condition and the associated privileged vocabulary hypothesis are false or incoherent.

III. INTENSION, EXTENSION, AND SYMBOLIC EXPRESSIONS.

INTENSION AND EXTENSION.

Let us go back to a passage quoted earlier;

> ... if a certain behaviour pattern (or input-output function) can be altered in a way that is rationally connected with the meaning of certain inputs (i.e. what they refer to as opposed to their physical properties alone), then the explanation of that function must appeal to operations upon symbolic operations ....[1]

The crucial point as far as Pylyshn is concerned is that behaviour alter in correspondence with a change in the 'meaning' of the input, for the class of changes in behaviour due to changes in input will include instances where the behavioural change (change in output) is due to non-symbolic, architectural functions. Hence we cannot say, for instance, that 'large' alterations in output must be due to symbolic processing and 'small' must be due to aspects of the functional architecture (unless we are prepared to define 'small' as less than 20 milliseconds).[2] Larger response durations could well be due to iterations of basic architectural functions. Neither can symbolic processes be identified by means of the correlations obtained between input and output. Strong correlations involving large changes in reaction time or type of response may be due to purely architectural considerations in split brain patients or people under the influence of drugs, whilst very small and subtle changes may be due to the nature of the alogarithmic processes in question, such as the differences in reaction time that might ensue when mental division is performed by the trial and error multiplication of the divisor, as opposed to dividing a column at a time and carrying the remainder.

---

[1] see p. 55 above.
[2] see Posner (1978).

Neither can we isolate symbolic psychological functions by an examination of the inputs involved. Initially it may seem that response variance in a study where the experimental stimuli are undifferentiated expanses of colour must be due to just the physical properties of the inputs, and thus mediated by non-symbolic processes. However, the experiment could be an investigation into the psychological reality of propositional logic, where each colour stands for a logical symbol. Similarly, an experiment utilising mathematical formulae or passages from a learned text could well be just a test of perceptual discrimination where the subject has to count the number of x's or vowels on the page as quickly as possible.

Inferences drawn about the nature of psychological processes from reaction timed data, or any other dependant variable, depend upon the validity of the ancillary hypotheses held. It is these hypotheses which shape the significance of the empirical data.[3] Reaction time is seen as a potentially fallible and indirect measure of underlying cognitive processes, to be interpreted in the light of other assumptions:

> Whereas, often, there is a correlation between the duration of a physical event and such purely alogarithmic properties as the number of steps taken, (and) the particular steps taken .... that is not always the case. There are certainly cases in which time differences arise from properties of the physical realisation that are unique to the particular occasion or instantiation ... and therefore are, in general, irrelevant to the alogarithmic, or process, explanation.

He continues:

> Using a computer as an example, we can see that some time differences might arise because a signal has farther to travel on a particular ... occasion because of the way the machine is wired or the way the alogarithm is implemented in it, or that some differences may arise from variable-delay effects unique to specific occasions. An example of the latter case is the delays caused by the distance a moveable arm must travel in making a disk access in some implementation and on certain

[3] Pylyshn (1984), p.126, 128.

> occasions, unrelated to the content of the memory or the alogarithm used ... Consequently ... measuring the times involved does not help us distinguish different candidate alogarithms ... time measurement alone cannot be taken as measurement of the alogarithmic process.[4]

Hence the important supposition that if behaviour alters in accord with the 'meaning' of the input, the input-output function is tapping alogarithmic or symbolic constituents of mental function, thus the appropriate explanation is cast in these terms. We would not for instance expect the amount of time it takes for a moveable arm to reach a particular point on a disk to depend upon the meaning of the input, although it may depend upon its length or intensity, - and the same can be said for the functional-architectural aspects of human memory.

We can rephrase and simplify the cognitive penetrability criterion like this:

> When behaviour alters in correspondence with an alteration in the meaning of the input, we can infer that the behavioural change is due to a symbolic/alogarithmic change (in state) as opposed to a non-symbolic architectural or physiological change.

The supposition underlying this must be that changes in the meaning of inputs will be mirrored by changes in symbolic state, and this independently of any change in output that may occur.

The first part of this statement reflects Pylyshn's contention that:

> symbolic codes ... reflect all the semantic distinctions necessary to make the behaviour correspond to the regularities that are stated in semantic terms.[5]

This is a different assertion from the following, however:

> symbolic codes ... reflect all the semantic distinctions *sufficient* to make the behaviour correspond ...

---

4 *ibid*, p.126, 127.

5 cf p.40 above.

This is the significance of Pylyshn's rejection of behaviourism. His position is that mental states bear only a contingent relationship with subsequent behaviour.. To deny this is to allow that mental states bear a logical or necessary relationship to subsequent behaviour in the fashion of Logical Behaviourism, a move which must deny representational psychology of its domain. The criteria for representationality entail that the way cognitive or representational processes unfold have a high degree of independence from the organism's causal interactions with the world. Mental states would become synonymous with disjunctions of behaviour in given stimulus situations (see chapter II). But this is what would be entailed by saying that a given symbolic (mental) state is sufficient for a certain behaviour to occur. This is equivalent to the assertion that 'if mental state x occurs, behaviour y must invariably ensue', a logical entailment which would enable us to "deduce the truths of mental ascriptions from the truth of behavioural ascriptions".[6] Mental states would become equivalent to the propensity to behave in such and such a way. Contrariwise, on the representational model of mind a particular psychological state such as the belief that the building is on fire, may lead to a diverse variety of behaviours which have nothing in common except that they all constitute ways of getting out of the building. This belief then, is only contingently related to any one of these behaviours. Given appropriate other conditions, such as the desire to immolate oneself, the same belief may lead to no behaviour at all. Thus, saying that a particular symbolic state is *necessary* to make behaviour correspond to certain semantic regularities is to say that, given certain initial conditions, a particular behaviour will ensue. But it remains allowable that the self same symbolic state will lead to a different behaviour, or to no behaviour at all. Consequently, we have the position, which is implicit in the concept of 'symbolic expression', that the identities of symbolic expressions are are not logically dependent upon the behaviours which they may cause or be conjoint causes of. They must therefore be determinate 'in the head' over and above dispositions to behave.

Given that the 'meanings' of internal symbolic states are not derivative of subsequent behaviour, (although behaviour is *evidence* for the postulation of the presence of this or that symbolic state in a given situation), the cognitive penetrability criterion relies upon a construal of

---

[6] cf p.28 above.

the relationship between cognitive input and process which ensures or allows that changes in the meanings of inputs are mirrored by changes in inner symbolic states or expressions. (The condition would serve no purpose if one could infer a change in symbolic state merely from the fact that a change in input leads to a change in output. This may happen and the change 'in the head' may be just in the functional architecture. Strictly speaking, the input-output function may be exactly the same in both symbolically mediated and non-symbolically mediated cases, thus the crucial step in the argument is the inference that if the meaning of the input has changed, then the cognitive processes involved are symbolic, which follows from the supposition that all relevant semantic distinctions are mirrored by symbolic distinctions. Succinctly, a change in the meaning of the input is a necessary condition for the ascription of underlying symbolic processes).

Of central significance then is the mirroring of semantic distinctions by internal symbolic states. In characterising these states Pylyshn allies himself with Fodor's language of thought hypothesis.[7] According to this view symbolic states are internal physically instantiated states which share many of the properties of natural language. If we view the internal medium of representation in this way many of the pre-theoretic requirements upon mental representation fall simply into place. Foremost amongst these is the requirement whereby:

> it is typically under an opaque construal that attributions of propositional attitudes to organisms enter into explanations of their behaviour.[8]

An opaque construal of a propositional attitude occurs when a belief that, desire for, intention to (X), (where X is a sentence like 'Mark Twain is a witty author') is attributed to someone in such a fashion that existential generalisation and/or substitutivity of identicals do not apply to the embedded attitude clause as in 'John believes that"Santa Claus has a red nose"'.[9] Consequently, in saying that John Smith believes that 'Santa Claus has a red nose', we are not committed to a), supposing that there is a real relationship between John Smith and Santa Claus, b) supposing that

---

[7] Pylyshn (1984), p.193-6. Fodor (1975).
[8] Fodor (1980), p.66.
[9] cf *ibid*, p.72.

there is such a thing as Santa Claus, or c) supposing that it must follow that John Smith believes that St Nicholas has a red nose. Thus in more typical situations wherein I may simply desire my neighbour's Porsche, this attitude does not consist in a relationship between me and the Porsche, - for my neighbour may not have a Porsche. Rather, my attitude is *something like* a cerebral inscription which reads 'want Porsche', and, just like a sentence of written English, seems to possess this sense of itself, intrinsically, as opposed to having its meaning by virtue of being in a specifiable relationship with an existent object, as would be implied by a Skinnerian analysis of language.[10]

So opacity allows us to liken propositional attitudes to inscriptions of mentalese, or sentences in the head. We may have different beliefs about the morning star and the evening star even though they are, in actual fact, the same star. I may believe that 'Mark Twain is a witty author' yet vigorously deny that 'Samuel Clemens is a witty author', in which case it seems appropriate to suppose that, in some way, I have a token of the former sentence in memory, but not of the latter, even though construed transparently as designating the same object the two sentences have the same meaning. We say that 'Mark Twain is a witty author' and 'Samuel Clemens is a witty author' have different intentions, but the same extension, for they both refer to the same person.

Pylyshn embraces the sentence analogy as we can see:

> the primary form in which the representations are expressed consists of discrete sentence-like symbolic expressions.[11]

Now, I want to argue that Pylyshn is subtly guilty of an old mistake. In its most primitive form the idea behind the mistake is the traditional supposition that the meaning of a word or sentence is, in its psychological aspect, a mental image or experience. This tactic bears certain parallels with the mental sentence strategy. In both cases it is assumed that the semantic content of a mental state is a discrete or atomic property of that state. This is like the idea that the content is a thing (like a mental picture) or a penumbra surrounding the arrangement of neurons or sequence of expressions that constitutes the embodiment of that state. It suggests that the content is intrinsic to that small part of the system at a given time.

---

[10] cf *Verbal Behaviour*, (1957).
[11] Pylyshn *op. cit.*, p.194.

A fanciful analogy is to take that bundle of neurons in my brain which (supposedly) correspond to my belief that 'flies are disgusting' and hold them steady in a 'spatial temporal stasis beam', then (somehow) change the nature of the world, change the causal history of the belief, change the organisation of the rest of my brain, and turn me into a creature which regards fly salad as haute cuisine. The idea we are considering is the suggestion that this belief of mine would remain unchanged by virtue of some immediate quality it possesses, regardless of its relations with other objects. This immediate quality is associated with what we informally think of as the 'meaning' of a sentence or belief, but which strictly speaking is its 'intension'. Bearing in mind that the 'extension' of a word or sentence is the set or domain of objects of which it is true (the extension of 'rabbit' is the set of rabbits), here is a passage from Putnam which clarifies these distinctions:

> ... consider the compound terms 'creature with a heart' and 'creature with a kidney'. Assuming that every creature with a heart possesses a kidney and vice versa, the extension of these two terms is exactly the same. But they obviously differ in meaning. Supposing that there is a sense of 'meaning' in which meaning = extension, there must be another sense of meaning of a term is not its extension but something else, say the 'concept' associated with the term. Let us call this 'something else' the intension of the term. The concept of a creature with a heart is clearly a different concept from the concept of a creature with a kidney. Thus the two terms have different intension. When we say that they have different 'meaning', meaning = intension.[12]

He goes on to note that 'Most traditional philosophers thought of concepts as something mental', and thus that meanings are mental entities. However, even those philosophers who rejected this picture, feeling that meanings must be public property and graspable by all, did not doubt that understanding a word (knowing its intension) was just a matter of being in a certain psychological state, that grasping meanings is an individual, autonomous psychological act.[13]

---

[12] Putnam (1975), p.217.
[13] cf Putnam *op. cit.*

The position we are trying to articulate then, at a first approximation, is the idea that the meanings or symbolic virtues of mental, representational states are intrinsic qualities of those states.

On the basis of 'creature with a heart' type examples the traditional view of meaning allowed that two terms might have the same extension yet differ in intension. However, it was presumed that two terms cannot have the same intension but differ in extension. But 'knowing an intension' is a matter of being in a certain psychological state, thus psychological state determines extension. But the extension of a term or sentence is what it refers to and is true of, thus psychological state, where we construe this as a state 'of the head', determines intrinsically what it refers to, what it is true of, and hence, broadly speaking, its semantics and content. In short, intension determines extension, therefore psychological state (qua state of the head) determines itself its content and what it refers to. [14]

A brief terminological digression: an 'autonomous' or 'narrow' psychological state is a state of a kind which presupposes the existence of no individual apart from the subject to whom the state is ascribed (hence they are opaque). Consequently narrow psychological states are, as it were, completely in the head (see p. 52 above).. Other psychological states are 'wide' psychological states. Thus if we construe my belief that 'Bob Jones is bald' in such a way that it essentially involves Bob Jones himself (an object external to me), it is a wide psychological state.

So to complete the above statement; traditionally, narrow psychological state determines intension, which determines extension, and thus determines itself its content and what it refers to.

Pylyshn's view is analogous to the 'traditional' view, and both views are incorrect:

Firstly; his 'symbolic expressions' are discrete sentence-like objects which are, implicitly, 'in the head'.

Secondly; they are construed opaquely, hence they are narrow. By construing 'Mark Twain is a witty author' as different from 'Samuel Clemens is a witty author' we can explain why these two 'thoughts' might cause different behaviours, even though transparently speaking they are the same 'thought'. Representational psychology of the type Pylyshn advocates is almost by definition the psychology of opaquely construed psychological states. Opaque construals of propositional attitudes are not

---

[14] cf *ibid*, p.221.

liable to existential generalisation, therefore no individual apart from the subject to whom they are ascribed is presupposed in their ascription.

Thirdly, just as in natural language two intensions may have the same extension (creature with a heart, creature with a kidney), so in Pylyshn's theory there may be 'synonymous' symbolic codes:

> There may be synonymous expressions - sets of codes with the same semantic content. Such codes might be functionally, but not semantically, distinguishable.[15]

Fourthly, however, and crucially, one intension, - one token symbolic expression, - cannot have two extensions, for, as in the traditional view, intension determines extension, and symbolic expression determines (its) semantics:

> symbolic codes ... reflect all the semantic distinctions necessary to make the behaviour correspond to the regularities that are stateable in semantic terms.[16]

This dichotomy is apparent in a passage already quoted. Bearing in mind that autonomous psychological state being 'completely in the head' will be determined by neurophysiology:

> Effects can penetrate upward through the levels, since each level is supervenient on levels below; that is , there can be no differences at level n unless there is some difference at level n-1, even though the converse is not true (because of the multiple-instantiation property of ascending levels; supervenience of psychological states on biological states entails that there cannot be two different thoughts unless there are some biological differences between the two underlying brain events).[17]

Cognitive penetrability follows from this supposition, for as long as a given symbolic expression can have only one extension, - refer

---

[15] above p.49.
[16] above p.40.
[17] above p.51.

determinately to only one class of things, - it follows that if 'a certain behaviour pattern ... can be altered in a way that is rationally connected with the meaning of certain inputs (i.e. what they refer to, as opposed to their physical properties alone)' then a symbolic change must be involved. That is, if reference changes, the symbolic code changes, hence by virtue of an observed change in the reference (meaning, semantics) of a given cognitive process we can infer that this process must involve the manipulation of inner symbolic codes rather than non-symbolic alterations of the functional architecture as supposed by the analogue theory of imagery.

As it happens though, it can be shown that intension, - narrow psychological state, - does not and can not determine its reference, and hence its semantics, and hence its 'meaning' roughly speaking. It therefore does not follow that an observed change in the 'meaning of certain inputs', (which as we have seen may or may not need to be accompanied by an alteration in behaviour), need be accompanied by any change in the underlying symbolic codes. The sense in which these codes possess meaning more or less intrinsically thus dissolves, which destroys the integrity of the symbol level, which in turn allows us to consider purely architectural, physiological or analogue structures as undertaking procedures previously thought of as essentially symbolic.

SYMBOLIC EXPRESSIONS.

As our currency is the subtle and arcane it is necessary to go back and underline the sense in which Pylyshn intends his notion of symbolic code to be understood. Whilst articulating this conception we will criticise it, removing barriers to the interpretation of psychological process which I will eventually espouse.

There are two senses in which cognition may be taken to require a level of articulated symbolic expressions: Firstly; it may be held that we require a level of symbolic expressions only in order to be able to state in a finite surveyable form, what a complex device such as a brain or a computer is doing. In this case the use of expressions is cognate only to the mode of description, the device or brain does not explicitly represent to itself the rules it is said to be following, they are only implicit in its operation. The knowledge the device or brain may exemplify in the

operations in question is then said to be procedural, as opposed to declarative. For instance , Dennet relates an instance where the designer of a chess playing program suggests that "It thinks that it should get its queen out early", yet as Dennet observes " ... for all the many levels of explicit representation to be found in that program, nowhere is anything roughly synonymous with 'I should get my queen out early' explicitly tokened". [18] In this particular case then the expression 'get queen out early' or similar is only a way of describing the system 'from the outside' rather than an explicit representation the system has and uses. There is nothing in the system that *is* this expression.

The second sense refers to those cases where there are expressions (in the language of Thought perhaps) which are explicitly tokened and used by the system. In this case the utility of expressions is not confined to or determined by the exigencies of description and explanation. In this case we would say that there is something in the system that is the expression 'that p', and hence is the representation for the system (as opposed to just the observer or describer) 'that p'.

Pylyshn's conception must be the second sense if our criticisms are to be pertinent, for unless symbolic expressions are determinate within the system it will naturally follow that a change in the meanings of the inputs, or any external change of circumstances relevant to the characterisation of internal 'symbolic states', will involve a change in those states much as suggested by the cognitive penetrability criteria. Internal states would necessarily mirror the external relations and circumstances which they are cited to explain. It is only if Pylyshn's symbolic codes are psychologically real in accord with the second sense above that it is an empirical question as to whether they mirror the semantics of their referents.

And this is how Pylyshn argues;

> to be in a certain representational state is to have a certain symbolic expression in some part of memory.

Immediately after making this claim he contrasts his position with that of Davidson and Geach, who allegedly:

> ... insist that it is unnecessary to assume that something (an internal property or a symbol) corresponds to P to explain

---

[18] Dennet (1978), p.107.

> how people can be in the state of 'believing P'. All we need, according to this view, are certain states of an organism which function in a particular way - namely, in a way correctly described (from the outside) as 'P - believing'. This is a kind of adverbial theory of intentional states which simply pairs functional states with belief (or other propositional attitude) descriptions without the step of positing any articulated substates or symbols that are the representations of P.[19]

Elsewhere (of cognition) "... we must view it as computing over symbols" where "the formal symbol structures mirror all relevant semantic distinctions".[20]

The idea of a token sentence inscription seems well suited to the role of an item which possesses meaning of itself, yet which is subject to certain syntactic or mechanical alterations to its form which may also change its meaning, hence Pylyshn's eagerness to characterise his symbols as articulated, sentence-like expressions.

There are many general reasons for assuming an explicit sentence-like medium of representation, the 'mental sentence' or 'language of thought' proto-model. The more informal of these I will discuss in the next chapter. I will now turn to a semi-technical rationale for this construal of representational symbols. This argument pertains quite closely to the theory of computation and so may appear to escape the criticisms I will develop below if not discussed separately. That is, whilst it will seem clear that something like the mental sentence which comprises the belief 'it is raining' (for example) cannot determine its extension, and thus does not have an intrinsic symbolic value (contrary to Pylyshn's theory), if the theory of abstract automata demands that potentially infinite behaviours (i.e. speech, counting) can only be accounted for computationally if recourse is made to a finite class of language-like expressions, - if symbolic expressions are computationally necessary,- then perhaps the former result is merely apparent.

---

19 Pylyshn (1984), p.29.
20 *ibid*, p.74.

FINITE STATE AUTOMATA AND TURING MACHINES.

Pylyshn insists that semantic interpretation is required if we are to be able to say what computation is being performed by a particular sequence of physical states within a computer or brain:

> to explain why the machine prints the numeral '5' when it is provided with the expression '(PLUS 2 3)' (with the symbols given their usual interpretation) we must refer to the meaning of the symbols in both the expression and the printout.[21]

Pylyshn makes a mistake of principle here. Whilst it is true that we must give some interpretation to the symbols in question if we are to be able to express the computation being performed, expressing the computation and interpreting the symbols are one and the same act, an act which is subsequent to or dependent upon the purely functional inter - relationships of the numerals as mere marks on paper. Strictly speaking we do not say 'ah ! PLUS, therefore it will print out 5', but 'it printed out 5, therefore the function is PLUS'. What confuses us is that this regularity is, in our culture, almost invariably referred to by the term 'plus' (or 'addition'), thus we tend to think that this word is associated with a nature of 'plusness' when in fact it is just a label for a type of regularity that may be found amongst mere physical states. If we were to visit some very isolated English speaking country where the word 'subtract' designated the function we refer to as 'plus', and the ten numerals we know were replaced by the first ten letters of the alphabet, our inability at first to interpret a computer's behaviour of printing out 'F' after being provided with '(SUBTRACT C D)' would not prevent the computer from performing additions, provided it maintained a consistent treatment (not 'interpretation') for the symbols it used and that the regularities in question were in point of fact all formally isomorphic with what we call 'addition'. In order to communicate what function the machine is performing to others, it is indeed necessary to interpret the computer's inputs and outputs, insofar as identifying it as a particular function also identifies the sort of roles the symbols in question will play. But this is not so much to explain the computation as to identify it. By citing the rules we

---

21 *op. cit.*,p.58.

do not thereby explain the rules. Having a vocabulary of identifiable functions like 'addition', 'subtraction', 'division' and so on is like knowing that such and such a card game is either canasta or bridge or old maid; if someone asks 'why did such and such an exchange of cards take place ?' we can explain this fact by identifying the game as bridge or canasta or whatever. But the situation in cognition is like witnessing a card game in a strange land where the pack has a variable number of cards, and none of the games played are familiar. Suppose that we invented a notation for recording the movements of the game, calling this movement 'a3', that movement 'b4ing' etc. In this case nothing could be explained by *interpreting the expressions* or *referring to the meaning of the symbols* until the rules of the game had been empirically determined and given names. In this way it becomes clear that in the sense of 'explain' that Pylyshn uses above, we can only explain what we already know.

More intriguing is the argument to the effect that the productivity of cognitive and computational systems, - that is, their ability to make, or be involved in, an unbounded number of distinctions, such as being able to utter a potentially infinite number of different sentences, - means that the regularities in question can only be captured if we regard the states as processing symbolic expressions:

> The regularities of a system with arbitrarily many functional states cannot be captured in a finite manner without some language - like combinatorial mechanism that ties together the systematic features of the set of state transitions ...

to which he adds;-

> ... indeed, this is a primary reason for a computational model appearing to be the appropriate one for cognitive science.[22]

The dichotomy here is the distinction between explaining the behaviour of a computer or brain in terms of its states, as opposed to explaining it in terms of the domain that the system operates upon,- and hence *representationally*. In automata theory this corresponds to the difference between a finite state automaton and a Turing machine. This is echoed by Pylyshn:

---

22 *ibid*, p.29.

> The difference between an extremely complex device characterised merely as proceeding through distinguishable states (but not processing symbols) and what I call a 'computer' is precisely the difference between a device viewed as a complex finite state automaton and one viewed as a variant of a Turing machine.[23]

FSA's are standardly described by either a state transition network or a machine table such as the one below:

	0	1
q^0	$q^1/0$	$q^1/0$
q^1	$q^0/1$	$q^2/0$
q^2	$q^2/0$	$q^0/1$

Input set X = (0,1).

Output set Y = (0,1).

State set Q = (q^0, q^1, q^2).[24]

Essentially then, an FSA is just a set of conditionals of the form 'If in state q1 and in receipt of input 1, produce output 0 and go to state q2', or generally, 'if in state 1 and in receipt of input x, produce output y and go to state 2'.

As we can see, the output and change of state must be mentioned explicitly for each input, thus a device capable of receiving potentially infinite number of inputs, as would be the case if it could, say, multiply any natural number by ten, would require a description in this format of a potentially infinite length. The machine table would, as it were, extend indefinately to the right as we included in the table the machine's response to each individual input. Likewise, if the device possessed a potentially infinite number of different responses to a given input - such as the ability to 'count' a string of any number of x's or y's, - it would require the explicit mention of a potentially infinite number of states. The

---

[23] *ibid*, p.70-1.
[24] cf Arbib (1969), p.57.

machine table would, as it were, extend infinitely downwards. In the FSA format of description any open ended or productive behaviour such as language production and comprehension, mental arithmetic, the ability to play chess - and much, if not most, of human cognitive behaviour is of this kind - will require for its performance a machine of infinite complexity.

A Turing machine (named after the mathematician Alan Turing) is an FSA connected to a reader/printer/mover device and a tape of potentially infinite length upon which is inscribed a series of symbols from a finite alphabet. The scanner reads a symbol at a time and either leaves it as it is or erases it and replaces it with another symbol from the alphabet. Then it either halts or moves to the left or the right along the tape. The symbol scanned and the state (or 'program') of the machine uniquely determines output, movement along the tape, and next state. Turing machines are generally considered to be an adequate formalisation of the notion of an effective procedure,[25] consequently any computable function is Turing computable. Thus in the context of the computational theory of mind the Turing machine is the proto-model for the explanation of thought. Given that Turing machines essentially involve a vocabulary of symbolic expressions the implication seems to be that human cognition must also make essential use of such a vocabulary.

The significance of the Turing machine for explanation lies in the addition to the FSA of specific strings of symbols on the tape and a reader/printer/mover device which alters these strings. The whole machine can thus be construed as manipulating expressions. Because the strings of symbols may be of arbitrary length the number of different inputs and outputs the machine can accept and make is potentially infinite even though the alphabet of symbols is finite. Because the alphabet is finite the machine's program, or set of states, can also remain finite, because we can define 'what happens' to each symbol, even if the computation goes on forever. And because we can construe the device as operating upon strings of symbols we can define the functions it performs recursively. The simple function mentioned on the previous page for instance can be defined as F(x) = T0, where T is any string of characters from the appropriate alphabet, in this case base ten integers. An FSA characterisation of the same function would in effect say '1-10, 2-20, 3-30,... 254-2540,... 7653467897-76534678970, ...' etc. for each input separately.

---

[25] This is Turing's hypothesis or Church's thesis, cf Arbib *op. cit.*, ch. 4.

However the necessity for expressing open ended functions in terms of operations upon 'symbolic expressions' does not establish that the function in question is used by the device except in a procrustean sense. For instance, consider a turing machine capable of a function which can be informally called 'doubling numbers'. Less informally we could say that the machine computes the function 'f(x) = 2x' in decimal notation. Now, it is tempting to infer from the valid conclusion that this function requires 'symbolic expressions' for its statement (for the above reasons) that the system must therefore 'think' *with* that expression, or a synonymous one. We think perhaps that if presented with a number to double we apply f(x) = 2x directly to it, so that the symbolic expression is not just a description of the computation in question, but the actual medium of the computation. This would appear to be the implication of statements such as:

> to capture the rule governed quality of computation the process must be viewed in terms of operations *on* formal expressions.[26]

and,

> It is only when the computer is described as *operating upon symbols* ... that we can explain its input - output behaviour in semantic terms ...[27]

(The idea that the symbolic values of symbolic expressions are the actual medium of thought and computation naturally goes hand in hand with the idea that token thoughts have intrinsic meanings or semantics.)

Here then is the program for a Turing machine that actually computes the function f(x) = 2x;

q^0 1 1 R q^1
q^0 2 2 R q^1
q^1 b 2 L q^2
q^2 1 2 L q^2
q^2 2 1 N q^3

---

[26] Pylyshn *op. cit.,,* p.69.
[27] *ibid,* p.72. His emphasis.

q^2 b b R q^4
q^4 2 b N q^3

Arbib (1969), p.131.

This machine operates upon base two numbers using the digits '1' and '2' instead of the usual '0' and '1'. It is a complete and obviously finite program although it can in principle multiply by two numbers of any length, provided they are in the appropriate notation and the machine is started on the rightmost digit. The terms q0 - q4 represent the machine's states , 1, 2, b the input - output alphabet (b = blank), and R, L, N, refer to the machine's movements along its tape, one space to the right, one space to the left, and no move respectively. We can translate, say, the third and fourth into informal English as follows; 'If I am in state q1 and currently inspecting a blank space, write a '2' in that space, move one space to the left, and go to state q2'; 'If I am in state q2 and currently inspecting the digit '1', erase it, replace it with a '2', move one space to the left, and remain in the same state'. (This sort of exposition isn't too inappropriate because in Turing's original 1936 paper he introduces the idea behind the abstract device in terms of a person with a pencil and paper who performs complex functions by going through a sequence of basic operations such as rubbing out one symbol and writing another in its place. A similar account is given by Rogers (1959)).

The firstly thing to notice is that nowhere in the program is there featured an expression synonymous with f(x) = 2x. Of course all seven instructions taken together exemplify f(x) = 2x, but at no point in the machine's operation will it ever follow this instruction, or even its base two analogue, f(x) = 21x.

Secondly, it is quite possible for a human 'computer' to mechanically follow the step by step instructions, multiply a binary number by two and yet be quite unaware of its decimal equivalent. In other words, one may be able to correctly perform a computation without having any cognisance of the symbolic significance of the tokens used, thus it follows that they are not used *as* symbolic expressions. Furthermore, there may be many possible symbolic interpretations of this computation viewed as a manipulation of strings of symbols. It could equally well represent a game, a decimal function that varies from f(x) = 2x at very high values of x, or a simple monetary transaction. We can note here an example due to Fodor (1978) where he raises the in-principle possibility that at the

machine language level, where no English terms are used, a computer chess program and a simulation of the Six Day war may in fact be indistinguishable when compiled.

So, in the case of the above device it is clear that no description of its behaviour more succinct than its actual program is actually used by the device and thus arguably is the medium of computation. Other renderings of its operation are molar, abbreviated descriptions and/or interpretations which may have an arbitrary number of synonymous and non-synonymous equivalents. Upon consideration this is exactly as we should expect, for the essential rationale of computation when performed by abstract automata is to break down complex 'intelligent' procedures into their most basic constituents and establish ways in which these elementary steps can be performed mechanically,- that is, without the intelligence that the task as a whole would seem to require:

> Let us imagine that the operations performed by the computer are split up into 'simple operations', which are so elementary that it is not easy to imagine them further divided.[28]

Thus, in a sense, the computational theory of mind (of which Pylyshn is obviously an adherent) is an atomistic theory of mental acts. Hence no intellectual act - nothing in the subject domain - will be achieved by mind or computer in the form in which it is presented for explanation.

Turing machines are creatures of mathematical theory and are rarely, if ever, actually constructed. The level of abstraction at which a Turing machine's instructions are written, that of a series of quintuples of monadic symbols , corresponds to the level at which the machine is conceived of as operating. That is, the machine 'does' the quintuples which are written for it. In more complex machines though instructions are given to the machine in a different vocabulary from that in which the basic operations of the computer qua physical artifact might be described. This is the distinction between machine language and programming language. Hence my argument above seems to imply that all computers which involve a programming language - which certainly includes all those used for cognitive simulation - do not in fact do what they are programmed to do ! But this is quite correct. One way to see this is to put

---

[28] Turing (1936), quoted in Arbib *op cit*, p.14.

the question 'If a given computer actually computes in its programming language, why then must a program first be compiled into machine language before it can be run ?' This point has been partially recognised in procedural semantics insofar as in that field it is thought that the comprehension of utterances in natural language is analogous to compiling (into machine language) and executing programs expressed in high level programming languages.[29] The implication is of course that we do not understand English in English, that the comprehension is actually effected at a more basic level. In this light it is natural to question the extent to which the 'symbolic expressions' Pylyshn adverts to inherently possess the 'meanings' or semantics which make them 'symbolic'. Which is the symbol level in our Turing machine example ? If it is at the level of the function f(x) = 2x or the informal expression 'doubling' then the symbol level is not the level at which machine or person actually computes - there are in fact other quite different alogarithms for performing 'doubling'.[30] This is not to say that this level is not symbolic, or is meaningless, or that this level is not absolutely necessary for the expression of regularities in the course of explanation, or that the terms used are not language - like 'expressions' (these are all points of Pylyshn's). Rather, in this and analogous cases these expressions are non-unique *descriptions* of what the system is actually doing; they denote and characterise certain regularities without comprising them. As such their metaphysical status is similar to that of emergent properties in Physics like 'heat'. There isn't really anything called 'heat', there is only molecular kinetic energy, but in having a commonly understood use and descriptive meaning 'heat' subsists and only hard line reductionists quibble. (This comparison is doubly relevant given that Pylyshn allies himself to Theoretical Explanation (in the full sense of the term) as used in Physics (chapter II above).

As modes of description the meanings attributed stem not in any essential way from the device in question, but in the expression having a generally known and accepted meaning in natural and theoretical language. That is, the meaning of 'doubling' is not derived from the set of computers and people it is attributed to as a 'symbolic expression' (as in 'it exists in peoples heads and then we discover it'), but in the general use of this word, like the general use of 'house', 'bird', 'run' etc.

---

[30] Arbib, *op cit*, p.131.

(It might be argued that because 'f(x) = 2x' is a formal expression it is the name for all those different alogarithms to which it is formally equivalent. The meaning of this expression then, unlike 'doubling' or 'P-K4', is a set of goings-on in brains and computers. But, although it is possible to program a Turing machine to perform this function in base ten notation, this program would again require a series of quintuples and wouldn't look anything like 'f(x) = 2x'. Although a program is in a sense formally equivalent to a succinct rendering of the function it is performing ('f(x) = 2x'), the machine does not use the latter expression, and that is the issue here. If we say that all these alogarithms - expressed perhaps as Turing machine programs - are doing the same thing and therefore they are intrinsically of such and such a type,- then 'same' here is always relative to some standard of comparison.)

IV. MEANINGS ARE NOT IN THE HEAD.

FOLK PSYCHOLOGY.

As we have seen, the traditional view of meaning is at most that meanings are mental entities, and at least that grasping meanings is an individual psychological act. These 'inner meanings' or meanings qua 'what is in the head' coincide with intensions, and are generally conceived to be determiners of extension. That is, something about whatever it is in my head that constitutes a thought about the Queen determines that it *is* a thought about the queen. The intension, it seems, has a kind of direction outwards towards its reference - what it is about and what it is true of - which we call its 'extension'. Although 'The Queen' refers uniquely to the Queen, a given extension may be referred to by more than one 'synonymous' intension. Hence 'the present monarch of the commonwealth', 'Charles Windsor's mother', and 'The head of the Church of England' are descriptions which express different intensions which share an extension. Traditionally it is thought that the converse does not occur; 'The Queen' does not refer to H.R.H. on some occasions and to nitric acid or the colour of snow on others.

Extension is tied to semantics insofar as the extension of a noun phrase is what it is true of, if it is true. Thus there is a sense in which two items which have different extensions will have different truth conditions, be true or false of different things, and so have different meanings.

This sort of account of meaning is implicit in that "loose network of largely tacit principles, platitudes and paradigms" which Stich and Churchland call "Folk Psychology",[1](and also in Pylyshn's conception of 'symbolic expressions' as we have seen). It is typified by, and exemplified in, such constructs as 'belief', 'desire', 'aim', 'fear', 'recollection', 'anticipation', 'knowing', 'expecting', and simple generalisations couched in terms of them like 'If x desires y and believes that doing z will enable him to attain y, then in general, and all things considered, x will do z'.

---

[1] Stich (1983), p.1. Churchland (1979).

Because beliefs desires and the like are in the normal course of conversation ascribed with sentences it is a natural extension of the folk paradigm to view these constructs qua 'intensions in the head' as akin to token sentence inscriptions. Thus we might speak of a 'snow is white' as being that 'snow is white' which is instantiated by a state of my brain at this point in time , just as we can mention different tokens of the phrase 'snow is white', of which this is the third in this chapter. We do not need to suppose, of course, that the propositional attitudes are realised in the brain in a form which might be read with a magnifying glass; only that the token phrases correspond in some way to instances of kinds of psychological state, where the precise nature of the correspondence is a philosophical and scientific question of no direct significance to the ordinary conception of these attitudes.

There are a number of reasons for this propriety. Foremost amongst these is the ease with which the almost common-sense idea of the 'mental sentence' is able to bridge the gaps between mental representation, the attribution of mental representations, and the 'ordinary' objects which these representations, in the main, refer to. What follows is closely based upon Stich (1983) p31-40. Briefly, using belief as an exemplar:

1) Beliefs are standardly attributed and named by linguistic constructions involving an embedded sentence, i.e. 'James believes "that p"'. The embedded sentence represents the most precise instrument available to the ordinary person for specifying the belief that James possesses.

2) The fact that beliefs and other attitudes ostensibly express two place attitudes is easily accounted for in ordinary language. We are able to distinguish James' belief from the object of that belief. A person may believe something which has no real object for instance, like 'unicorns are white'. People may have different attitudes with respect to the same object: Mary may believe that there is a god, whilst John may hope that there is a god.

3) The semantic properties of beliefs, their objects, and the sentences embedded in belief ascriptions all coincide. That is, if my belief that the world is round is true, the content sentence 'the world is round' is true, and the world is in fact round. It would be nonsense, for instance, to claim that John's belief is true, though what he believes is false. If, again, my belief entails John's belief, (I believe that the prime minister is a man, he believes that the prime minister is mortal), then the content sentence

(embedded sentence) for my belief entails the content sentence for his belief.

4) If, following Stich, we call sentences of the form 'S believes "that p"' belief sentences and the embedded sentence p the content sentence, then we can note that in both beliefs and in ordinary language, the semantic properties of belief sentences are independent of the semantic properties of content sentences. Hence it may be that S believes 'that p', but 'that p' is false (James genuinely believes that the moon is an alien spacecraft). Alternatively, it may be false that S believes 'that p', whilst 'that p' is true (Ceaucesceau believes that Stalinism is a repressive political system). The truth value of the content sentence tells us nothing about the truth value of the compound. We cannot conclude from the truth of 'that p' that 'James believes "that p"' is either true or false. Also from the truth of 'S believes "that p"' and p entails q, we cannot infer the truth of 'S believes "that q"'.

This is related to the opacity and non-extensionality of beliefs and belief sentences. Because one extension may be referred to by a number of intensions a person may be privy to one description of a person or object without being privy to others, consequently there is an equivocation in the senses of belief sentences as illustrated by this example;

S believes that Fa	(i.e. Frank believes that David is a communist)
a = b	(David is the university proctor)
S believes that Fb	(Frank believes that the university proctor is a communist)

This sort of inference is not generally valid. This relates to Brentano's thesis that 'intentional inexistence' is the mark of the mental (Brentano (1874)). This is the idea that mental phenomena are characterised by an immanent objectivity, an inclusion of an object - the thing thought of - "that is short of actuality but more than nothingness"[2], unlike the simple actuality of physical relations. Thus 'John is thinking about a sports car' does not imply that there is a sports car, like 'John is driving a sports car'. Consequently, failure of existential generalisation and referential opacity are seen as important indicators of mental phenomenon.

---

[2] Encyclopedia of Philosophy, p. 201.

5) The logical relationships of beliefs mirror the logical relationships of their content sentences. For instance, if I believe that all a's are b, and that all b's are c, I will typically, though not invariably, come to believe that all a's are c. This provides folk psychology with a ready-made network of rough generalisations pertaining to content sentences to serve as a framework for subsequent folk generalisations pertaining to beliefs, goals, desires etc.

This marriage of language and folk art gives a theory-like presence to Folk Psychology. Insofar as beliefs, desires etc. exist primarily within this theory there becomes something resembling a truth of the matter in cases of folk-psychological ascription and explanation. What common sense and intuition tell us becomes authoritative when the issues and constructs in question are by their very nature 'common-sense'. If there is a scientific sense of belief (e.g.) then it is not *synonymous* (except accidentally) with a concept that each generation has learnt at their mothers' knee. The scientific correlates of belief may or may not have much to do with the ordinary person's concept of belief, but even if science should supercede folk psychology, it has no power to deny that within folk psychology rules and concepts take a certain form. It is not always appreciated that the common-sense framework of the world, Sellars' (1963) 'Manifest Image', has an existence that is substantially independent of scientific fact. Most of us know nowadays that 'solid' objects are not really solid; as Eddington would say, 'Physics tells me that my desk is mostly empty space', but this does not show that when the ordinary man says 'solid' he really means 'mostly empty space'. By the same token, when we use terms like 'belief', 'aim', 'desire', etc. in non-scientific contexts, we do not 'really' mean 'symbolic state number 3289'. Thus folk psychology and common sense have their own standard of psychology, to which scientific psychology may either rise, oppose, or reconcile itself with.

Some matters of definition; 'autonomous' psychological states are those that would be shared by a person and his atom-for-atom replica. This follows intuitively from our previous discussions of 'narrow' and 'autonomous' states. Stich (1983) presents autonomy of psychological state as a fundamental regulative principle for a genuinely explanatory psychology. This serves to eliminate a number of erstwhile psychological

properties, such as 'remembering my fifth birthday', for my replica cannot remember *my* fifth birthday, although he may seem to remember it.

A principled way of putting the principle of autonomy is in terms of the supervenience of properties. Stich borrows from Kim (1978) here saying:

> The family S of properties supervenes on the family W of properties (with respect to domain D of objects) just in case, necessarily, any two objects in D which share all properties in W will also share all properties in S.[3]

So autonomous psychological states are those states which supervene upon the current internal physical properties and relations of the organism, and the Principle of Autonomy is the principle that psychological theories should confine themselves to these states. But if this is true Folk Psychology and its attendant propositional attitudes seem to be invalidated, for beliefs, hopes, desires and so forth seem to be intrinsically associated with their objects in the world, the things believed, hoped for and desired.

From this it is clear that 'formal', 'syntactic', 'functional' and 'computational' states of the organism are also autonomous in Pylyshn's model, for they are instantiated in the biological architecture, and thus supervene on the physical state of the organism ("supervenience of psychological states on biological states entails...", p.51 above). ('Syntactic' and 'formal' are two ways of saying the same thing; both are specified without reference to such things as meaning, semantics, reference etc.).

We have seen that the broad thrust of Pylyshn's position is that semantic level (belief and desire) generalisations are to be explained in terms of formal 'symbol level' regularities. Given that the former are, by definition, contentful and the latter syntactic, this makes Pylyshn an adherent of what Stich (1983) calls 'The Weak Representational Theory of Mind', which holds both that;

> the generalisations of cognitive science will be purely formal, applying to mental states not in virtue of their semantic properties, but rather in virtue of their syntax.

---

[3] Stich (1978), p.575.

and;

> semantic features are correlated with the syntactic type of the token

(the 'correlation thesis'), that is;

> mental states which are 'functionally identical' ... must have the same content.[4]

For, as we have seen, Pylyshn asserts that;

> each level is supervenient on levels below (p. 51)

and;

> ... differences in content always result in functional differences(p51).

The last quote does not imply that functional (syntactic) differences always result in differences in content, but this is not necessary for my argument; there may be different functional states with the same content. The counter-thesis to the correlation thesis is the supposition that a given formal, functional, syntactic or 'symbolic' state, insofar as it is supervenient upon the physical properties of the organism, is associated with, or has, more than one content.

The correlation thesis clearly underlies the cognitive penetrability criterion: It is only on the supposition that each functional or syntactic state type maps onto a separate content (has a separate content) that we can infer a change in autonomous state from a change in content. Content being what it is, a generic term for that which is possessed by states or objects to which truth conditions may be applied, the implication is that autonomous psychological states must alter in concert with their external objects. A change in the 'meaning' of an input is deemed a sufficient condition for change in narrow psychological state (as long as the input is apprehended, i.e. as long as it is not just a possible but an actual input to

---

[4] All Stich (1983), p.185-6.

the system), for the correlation thesis rules out the possibility of a given state having more than one content.[5]

Since Putnam's landmark paper 'The Meaning of "Meaning"' in 1975 a number of examples and 'thought experiments' have been described in the literature which conflict with the two central components of the traditional view of meaning which Pylyshn uncritically accepts in his cognitive theory, via the influence of the ubiquitous mental sentence proto-model. It is natural for a theory informed by this model to accept a), that each token 'symbolic expression' has its own determinate and intrinsic content but b), only one such content. The first thesis is required by the necessity that the codes have a symbolic value if this quality is to be cited as a theoretical quantity in explanation. The second follows from this on the grounds that this quality must be the same quality (for each token expression) on separate occasions of its use if it is to have the generality required of an explanatory construct. Both of these contentions are false: there exist well formed and normally ascribed psychological states which have no clear or intrinsic content, and equally well formed states to which may be attributed more than one content. These examples violate the correlation thesis and the traditional 'mental sentence' proto-model of meaning, and by so doing strip the cognitive penetrability principle of its force, and place Scientific Psychology in opposition with Folk Psychology to the extent to which the former seeks formal, alogarithm-based explanations of behaviour.

---

[5] There is an element of vagueness here: The canons of anti-behaviourism ensure that cognitive changes may take place in the absence of any overt change in behaviour - I may perform image manipulations in the absence of any eliciting stimuli, and emit no subsequent response upon completing my manipulations. There is clearly a sense in which the system may cognise or register data without the necessity of prior or subsequent stimuli or response. Informally speaking, there is a difference between noticing a pun on words (say) and not noticing it, even when initial conditions are the same and no change in behaviour ensues in both cases. This is inherent in Pylyshn's system for the symbolic codes reflect all the semantic distinctions *necessary* to make the behaviour correspond to the regularities that are stated in semantic (belief and desire) terms, consequently there may be symbolic changes which are unaccompanied by any change in externally ascribed semantic state, such as would occur if there had been no change in overt behaviour. The prospect of cognitive change in the absence of behaviour is not explicitly recognised by Pylyshn, but it is clearly consistent with his theory, and this possibility is all that my argument requires.

PSYCHOLOGICAL STATES WITH NO CLEAR CONTENT.

These may be divided into three categories. The rationale for discussing each in terms of 'conceptual schemes' will become apparent as we work through the examples, most of which are taken from Stich (1983). In each case we will take belief as an exemplar, but all examples may be generalised to hopes, desires, goals, intentions etc. We will assume, for the sake of the argument, that belief states are instantiated as token 'mental sentences'. The criteria for possessing a certain belief state then is the possession of a token of the appropriate content sentence in the appropriate. memory register. As one discrete token is sufficient for possession of a belief state, a simple avowal or utterance of the relevant sentence should be sufficient evidence for its ascription.

It is not necessary for us to suppose that there is a one-to-one relationship between sentences of English and symbolic expressions for the argument to be relevant to Pylyshn's theory. It follows that if one symbolic expression possesses a unique and intrinsic correlative symbolic value, then a set of n symbolic expressions must also possess a unique symbolic value or content. If there are any higher order group effects then these are ipso facto not accounted for in terms of the association of component symbolic states. If it is Pylyshn's idea that the symbolic codes *explain* beliefs, then there can be no indeterminism here.[6]

If it does seem that the content of a given belief state qua content sentence ('snow is white') varies from one environmental situation to another (as I will argue) Pylyshn could perhaps say that it is not the symbolic values of the expressions which have changed, but the differential individuation of those expressions involved on different

---

[6] This is to skirt very briefly over some profound issues. Douglas Hofstadter in his *Godel, Escher, Bach* (1980) alludes to a sense, derived from Godel's incompleteness theorem, in which a complex formal system, such as a brain may perhaps be, may have molar properties which amount to more than the sum of its parts, but which are not as it were 'applied from the outside'. (There may be theorems of a formal system which cannot be proved in that system). Perhaps there is a route here for Pylyshn to explicate some sense in which a given symbolic expression or set of such codes may possess both a 'basic' and an emergent content, and so escape the criticisms below. These considerations are however far beyond the usual ken of computational explanation in psychology, and are nowhere mentioned in Pylyshn's (1984), his central work, thus if there is any escape route here it is incumbent upon Pylyshn to explicitly formulate the manner in which this might be achieved. Having said this, I doubt that there is any way out for him in this direction, for he explicitly associates changes in content with functional, and hence physical, changes,(see previous page above) whilst the levels of meaning envisaged by Hofstadter are not created by any physical change, but are inherent within complex structures.

occasions - the difference being that 'snow is white' = expressions (a,b,c,d) on occasion one, and (b,c,d,e) on occasion two. But if the semantic level belief sentence remains the same, and each level is, by hypothesis, supervenient upon the level below, then we must be talking about the same physiological structure on both occasions; hence this tactic may succeed only at the cost of discarding a material basis for symbol level structures.

Impoverished conceptual schemes.

'Mrs T' according to Stich was a person who was at one time employed by his family. She was well over 80 at the time and remembered the assassination of the U.S. president William McKinley in 1901, an event which deeply shocked her at the time. If asked for instance 'was president McKinley assassinated ?' she would reply 'yes', and thus could be attributed a psychological state, or symbolic expression, or set of symbolic expressions, with the content 'McKinley was assassinated' by any reasonable criteria. In fact we can simply stipulate evidential considerations aside and say ex hypothesis that Mrs T has the appropriate sentence-like symbolic expression(s) in the appropriate memory location, sufficient, on the account we are investigating, for her to be imputed the content sentence 'McKinley was assassinated'. However, in the later years of her life her memory began to fade, and shortly before her death Stich recorded a conversation with her which went something like this;

S; Mrs T, tell me, what happened to McKinley ?
T; Oh, McKinley was assassinated.
S; Where is he now ?
T; I don't know.
S; I mean, is he alive or dead ?
T; Who ?
S; McKinley ?
T; You know, I just don't remember.
S; What is an assassination ?
T; I don't know.
S; Does an assassinated person die ?
T; I used to know that, but I just don't remember now.
S; Do you remember what dying is ?

T; No.
S; Can you tell me whether you have died ?
T; No, I just don't remember what that it.
S; But you do remember what happened to McKinley ?
T; Oh yes, he was assassinated.[7]

This example demonstrates the holism of belief. That is, in order to have a certain belief one must necessarily also have certain other beliefs (desires, hopes etc.) and attitudes. Davidson remarks:

> Beliefs and desires issue in behaviour only as modified and mediated by further beliefs and desires, attitudes and attendings, without limit.[8]

To believe something is never to believe one thing, but to apprehend a conceptual scheme. To paraphrase Wittgenstein, that one can never believe just one thing 'is not a more or less arbitrary point of departure ... but belongs to the very essence of what we call a belief'. Specifically:

> If we believe something at all it is not a single fact or a single proposition, but a whole system of propositions.[9]

Possessing a certain belief cannot consist in merely having a particular mental sentence in the appropriate memory register, as is clear from the example above. Although she avows the sentence 'McKinley was assassinated' she has no conception of what assassination or death are, and thus can hardly be credited with the belief-sentence 'McKinley was assassinated' as we would understand it. Similarly, whatever symbolic expression or set of such expressions underlies Mrs T's belief, they do not appear to possess of themselves the content 'McKinley was assassinated'.

To possess a belief one must have (at least) the appropriate sentence in the appropriate register *and* a network of further beliefs which interpret the first belief. The simple avowal of a single sentence like 'the house is on fire' is sufficient in normal circumstances for the correct imputation of a belief because of the ideological similarity of the population in question:

---

[7] Stich, *op. cit.*, p.55.
[8] Davidson (1970), in Honderich and Burnyeat (eds.), p.228.
[9] Wittgenstein, *On Certainty*, quoted in Brand (1979), p.9.

their concepts are similar; everyone knows what a house is, what a fire is, and so on. But this is an entirely normative fact about a population, a fact which tells us less about the nature of thought and belief than it does about the objects of thought and belief.

Suppose that at some future date Science develops a new set of concepts and technical terms. Suppose that the following sentence represents a deep strand of this new doctrine:

> Hydrogen atoms are single-petalled superheterodyning negentropy flowers.[10]

Suppose that a contemporary individual called Paul somehow comes to possess a mental sentence of this form. Thematically, it would be as if both Paul and certain future scientists all had a mental sentence of this shape - these English words and letters - inscribed on some internal belief-screen. By concentrating on 'shape' rather than 'content' or 'meaning' we are limiting ourselves to psychologically autonomous, syntactic, formal or functional states, so by the lights of Pylyshn and the WRTM Paul and the future scientists have, in this mental sentence or symbolic structure, a belief with the same content. But this is scarcely credible. Even if Paul comes to avow this sentence he will not, as it were, 'know what it means', for the theory from which it comes is unheard of today. A similar situation would be engendered by an eight year old saying '$E = MC^2$'. Unless the child is a genius his mere possession of this equation as a belief token will not confer upon him an understanding of the General Theory of Relativity.

One more example: Stich reports that his six year old knows that the Star of David has 'six' points, however her command of addition is such that she is not aware of what six plus three is.[11]Can we adequately attribute to her any belief involving the number six ? Exactly the same circumstances could be true of a mentally retarded or senile adult, stroke victim or psychotic; anyone whose conceptual scheme or 'world view' is impaired or incomplete compared with the norm. Because beliefs and propositional attitudes are attributed with tokens of a shared language there is a normative element which, insofar as it ensures a commonly understood meaning for the median 'snow is white' (say), will fail to

---

[10] Stich, *op. cit.*, p.57.
[11] *ibid.*, p.143.

render any clear or intelligible content for unusual cases such as those cited above. Deprived of a complete context the possession of the autonomous correlate of a single token inscription is not sufficient for the attribution of the conceptual content that is normally associated with that inscription.

Idiosyncratic conceptual schemes.

There is a body of literature on the subject of belief perseverance. which suggests that if a subject is duped into believing a spurious fact about him or her-self in an experimental situation, he or she will continue to exhibit a residual belief in the putative fact after being informed that his/her results in the psychometric test or personality measure were in fact faked.[12] This I mention only to indicate a kind of non folk-psychological, counter-intuitive, and hence scientifically interesting generalisation which personal beliefs of the form 'I am extremely intelligent' , 'I have leadership qualities' and 'I have homosexual tendencies' might be involved in.

Suppose someone comes to possess a belief we would attribute with the content sentence 'I have homosexual tendencies'. In most cases this person will share with his contemporary English speakers of sound mind and disposition a substantial number of beliefs about sex and homosexuality, such that tokens of this belief will mean much the same thing when attributed to members of the community. But being an adult co-linguist of sound mind does not ensure a unanimity of sense amongst all possessors of belief tokens of a single type. Suppose that our subject, quite sanely, possesses an idiosyncratic understanding of the notion of sex as follows:

> What sex a person is, is not a function of anatomy. Maleness and femaleness are basic, irreducible properties of people. These properties are often correlated with anatomical differences, but sometimes they are not ... It is no easy matter to determine what sex a person is. You have to know quite a lot about their personalities, their goals and aspirations, the way they interact with other people etc ...[13]

---

[12] Ross, Lepper and Hubbard (1975), Ross (1977).
[13] Stich, *op. cit.*,p.138.

If we convince this person that he has 'homosexual tendencies' in a belief perseverance experiment, the belief which he comes to hold will be quite different from that which most of his community would come to hold in the same situation. He may, for instance, judge as 'homosexual' the relationship between Jane and John, whilst viewing David and John's relationship as perfectly heterosexual. But how then can we characterise this person's residual belief as 'I have homosexual tendencies' ? And to the extent that this characterisation of his belief state is indeterminate, any substitution of this content sentence in scientific generalisations which include sentences of the form 'S believes "that p"' will render these generalisations similarly indeterminate.

This point is generalizable to any concept or subject matter about which we have something resembling a theory. Is my concept of 'introversion' the same a yours ? Quite possibly not. If certain people may have different concepts of homosexuality or free enterprise then others may have different concepts of bread, walking, or furniture. The mere possession of a discrete belief sentence or set of 'symbolic expressions' will not confer a similarly discrete or univocal content.

Incoherent conceptual schemes.

> Suppose that ... a Mr Binh, is a recent immigrant to the United States whose mastery of English is rather shakey. A bright and attentive man, Binh is anxious to learn as much as possible about his adopted country. On his first day off the plane he overhears a conversation about a Mr Jefferson, whose exploits are of obvious interest to the people on whom he is evesdropping. Unknown to Binh, the people whose conversation he overhears are avid TV fans, and they are discussing the most recent travails of (a) fictional black dry-cleaning magnate ... The next day Binh begins citizenship classes and he hears that Jefferson was a statesman, an inventor, and a major figure in the early history of America ... On the third day Binh hears some discussion of a Mr Feferman, a brilliant logician. However, with his ear not well attuned to spoken English, Binh hears 'Feferman' as

> 'Jefferson'. Finally, on the fourth day, Binh meets an old friend and has a long chat about what he has learned of his new country. 'I am' he says, 'very anxious to learn more about this fascinating fellow Jefferson, the black patriot and statesman who made significant contributions to logic whilst building a dry cleaning empire'.[14]

With confusions of reference we have a type of ostensibly well formed sentence-like mental state which admits of no intelligible content. 'Jefferson, the black dry cleaning magnate ...' because it conflates the identities of three people, either fails to refer or has no clear extension. If intension determines extension then, ipso facto, it has no clear intension, and thus no clear 'symbolic value' if it corresponds to a 'symbolic expression' or set of 'symbolic expressions'. Examples of this type may well be common in everyday life. It is easy to conceive of confusing Old Man's Beard and Ivy, Madras street and Barbados street, primary and secondary qualities, the victor of Austerlitz and the father of Napoleon III.

AUTONOMOUS PSYCHOLOGICAL STATES WHICH ADMIT OF MORE THAN ONE CONTENT.

This is the area in which the traditional conception of meaning breaks down most explicitly. Consider these examples:

> Tom is a contemporary of ours, a young man with little interest in politics or history. From time to time he has heard bits of information about Dwight David Eisenhower ... Let us ... assume That each time Tom heard something about Eisenhower, Eisenhower was referred to as 'Ike'. Tom knows that this must be a nickname of some sort, but he has no idea what the man's full name might be, and doesn't very much care. Being little interested in such matters, Tom remembers only a fraction of what he has heard about Ike: that he played golf a lot; that he is no longer alive; that he had a penchant for malapropisms; and perhaps another half dozen other

---

[14] *op. cit.,,,* p.145-46.

> facts. He has no memory of when or where he heard these facts, nor from whom.

However:

> Dick ... is a young man in Victorian England. Like Tom, he is bored by politics and history. Dick has heard some anecdotes about a certain Victorian public figure, Reginald Angell-James, who, for reasons that history does not record, was generally called 'Ike'. And ... in all the stories that Dick has heard about Angell-James, the gentleman was referred to as 'Ike'. Angell-James and Eisenhower led very different careers in different places and times. However, there were some similarities between the two men. In particular, both were involved in politics and the military, both liked to play golf, and both had a penchant for malapropisms. Moreover, by a quirk of fate, the few facts Dick remembers about Angell-James coincide with the few facts Tom remembers about Eisenhower. What is more, of course, Dick would report these facts using the very same sentences that Tom would use, since the only name Dick knows for Angell-James is 'Ike'.[15]

We can then suppose that the mental sentences or symbolic codes underlying Tom and Dick's beliefs are identical autonomous psychological states. But they do not have the same content. For instance, the belief that 'Ike liked to play golf' may be true of Eisenhower but false of Angell-James. We feel strongly inclined to say that Tom and Dick's beliefs are about different things, and so will possess different truth conditions. But if they possess different truth conditions they cannot possibly be the same belief. As semantics just is the question of truth and truth values their autonomous states have different semantics, and so are different representations.

Any suggestion that there must be some causal difference which distinguishes the beliefs of Tom and Dick is difficult to motivate, for neither has either met or seen 'Ike', and their immediate contacts with written or verbal information about 'Ike' can be stipulated to be the same.

---

[15] op. cit.,, p.60-61.

An analogous situation is found in another example due to Stich: Apparently the vegetables known as 'chicory' and 'endive' in America are known in the U.K. as 'endive' and 'chicory' respectively. Suppose then that we have an Englishman, Robin, and an American, John, who who both dislike vegetables and are woefully ignorant of the appearance of all but the most common species. Both are invited to a dinner party in their respective home towns. Both have heard, whether truthfully or falsely, that 'chicory' is bitter. Both are asked if they would like a 'chicory' salad and both tactlessly refuse saying 'no thanks, chicory is bitter'. Now, even if we stipulate that by hypothesis the symbolic states underlying the beliefs expressed by these two men are identical, it seems doubtful that this identity of autonomous psychological state ensures identity of belief content. Suppose that the vegetable the Americans call 'endive' is bitter whilst 'chicory' is in fact sweet. In this case John's belief is false and Robin's belief is true, and they surely can't be the same belief if they have different truth conditions. Here the content of belief states seems to depend upon the proximity of the subject and object of the belief, for if Robin and John were attending their respective dinner-parties at the same time and were teleported into each others seats just prior to being offered the salad, John's belief would become true and Robin's belief would become false. But at what point do their beliefs change in truth value, when John is halfway across the Atlantic, or when he swoops in over the Shetland islands ?

A similar phenomenon is exhibited by indexicals. The ubiquitous mental sentence proto-model suggests that if both you and I believe that the sky is blue then we both possess a token of 'the sky is blue' in mentalese in some mental register. However, if two different people both possess a belief or attitude which contains an indexical expression like 'I', 'you' or 'this', the content of the sentence in question will change depending upon who instantiates it. If I am being attacked by a bear and you are observing then 'I am being attacked by a bear' does not adequately characterise the content of both of our beliefs. Mental sentence identity then, does not confer content identity. Ipso facto, the 'symbolic expression' account which seeks to correlate with and explain beliefs qua 'semantic level regularities' does not possess the resources to account for variation in belief contents except at the cost of allowing variations in 'symbolic values'.

Stich makes the same point in his 1978 paper: Suppose that it is possible to create atom-for-atom replicas of human beings. Physical replicahood will then confer identity of autonomous, and hence syntactic, formal and/or functional states. Suppose that a replica of Stich has just been created. Stich believes that he has tasted a bottle of Chateau d'Yquem 1962. The embodiment of this belief in 'symbolic expressions' will be amongst the causes of his behaviour, such as remarking 'I have tasted a bottle of Chateau d'Yquem 1962', and being able to infer that 'I have tasted a French wine vinted in the 1960's '. If we ask Stich's replica if he has tasted a bottle of Chateau d'Yquem 1962 he will answer affirmatively, and so we may feel that this belief is amongst the causes of his utterance. Yet this cannot be the same belief that Stich possesses for Stich's belief is true whilst Stich's replica's belief is false, for Stich's replica has, by hypothesis, only just been created. If the alogarithmic embodiments of this belief are identical in both cases and lead to identical behaviours, they do not achieve this effect by virtue of their *content*. [16]

With the aid of conceptual tools borrowed from Science Fiction, the scope of this kind of example is completely general. Moving from self referential belief to beliefs about one's surroundings, we can imagine that in a time when the art of cryogenics has been perfected a person with certain true belief about his surroundings might be quick frozen, transported to another location and kept on ice for a century or two. [17] From our previous discussions it is clear that this person's narrow, autonomous, symbolic states will remain constant from the moment he is frozen to the moment he is unfrozen. Just prior to being frozen he may believe that his car is parked outside, there are many strawberry farms nearby, the weather is good and so on, and each of these beliefs has the potential to influence subsequent action. But upon defrosting his beliefs may be false, whilst the symbol level structures which are supposed to account for the content of these beliefs remain unchanged. Beliefs cannot be assimilated to these autonomous structures, hence beliefs cannot be 'in the head'. Any number of our beliefs about our spatial and temporal surroundings may be slotted into examples of this kind.

Similarly for beliefs about other persons: Suppose that I have a doppelganger on a planet far away that is virtually identical to Earth. This planet is so similar to Earth that there is a country there which goes by the

---

[16] Stich (1978), p.580.
[17] *ibid.*

name of 'New Zealand', and an ex-prime minister called 'David Lange' who is resident in this country. Now I believe that David Lange has five fingers on his left hand (four fingers and a thumb), and my doppelganger, being an atom for atom replica of myself, will exemplify all of the autonomous psychological properties that I do, and prima facie will have the same beliefs that I do. But David Lange on this far off planet has *six* fingers on his left hand, hence my belief is true and my doppelganger's false. The conclusion is familiar.

Which brings us to natural kinds and the thought experiment which initiated the debate in 1975. In 'The Meaning of "Meaning"' Putnam asks us to suppose that somewhere in the galaxy there is a planet called 'Twin Earth', which is exactly like Earth down to and including an atom-for-atom replica of each of us. We can also suppose that on Twin Earth they speak 'English', which is identical to our English. The only difference between Earth and Twin Earth is that the liquid referred to as 'water' on Twin Earth - the liquid which fills lakes and rivers, quenches thirst, falls as rain and so forth - is not H_2O but another substance which has a very long and involved chemical formula which we can summarise as 'XYZ'. Furthermore, XYZ is indistinguishable from H_2O in all but the most esoteric laboratory circumstances. Now, were a spaceship from Earth to visit Twin Earth, and were its occupants to learn that the local 'water' is in fact XYZ they would be inclined to say:

> On Twin Earth the word 'water' means XYZ.

Symmetrically, if a spaceship from Twin Earth visited Earth they would be inclined to say that:

> On Earth the word 'water' means H_2O.[18]

In other words, the sense, concept or intension of the word 'water' has the extension H_2O on Earth and XYZ on Twin Earth. From our perspective what the a Twin Earthians call 'water' simply isn't water, it is something else. At this point in the argument it may still be possible to maintain that there is something in the individual's concept of water which 'controls for' water's idiosyncratic reference to XYZ on Twin Earth, for by hypothesis, at least some people are aware that 'water' refers to two

---

[18] Putnam (1975), p.223-224.

different things on these two different planets. Putnam's point though, is that no-one on either planet need be aware of this variation in extension for it to be the case. If we went back to a time prior to the development of chemistry or the atomic theory of matter on Earth and Twin Earth no-one would know that Twin Earth water was undetectably different from Earth water. In particular, my doppelganger on Twin Earth could be my exact duplicate in feelings, interior monologues, thoughts and psychological states insofar as these are constituted by my physical, bodily and behavioural state or history - i.e. at least my functional architecture and symbol level. Thus there can be no differences in our concepts or intentions or symbolic states, nor between any pair of duplicates on our respective planets, yet we still each mean something different by the word 'water':

> Thus the extension of the term 'water' (and, in fact, its 'meaning' in the intuitive pre-analytical usage of that term) is not a function of the psychological state of the speaker by itself.[19]

Although my doppelganger and I are both in the same autonomous psychological state, our 'thoughts' have a different content. Our symbolic states are exemplifying neither an intrinsic and determinate symbolic value (there could be hundreds of Twin Earths each with a different substance going under the name of 'water', leading to there being as many different contents for the symbolic expression concerned as I have doppelgangers) nor the same symbolic value in different circumstances.

Tyler Burge makes an analogous point by allowing extension to vary as a function of understanding, within a linguistic community. In a by now familiar fashion he works from the premise that:

> On any systematic theory, differences in the extension - the actual denotation, referent, or application - of counterpart expressions in that-clauses will be semantically represented, and will, in our terms, make for a difference in content.[20]

---

[19] op. cit., p.225.
[20] Burge (1979), p.75.

A second premise is that, as we have seen, it is is specifically oblique construals of propositional attitudes which serve to characterise the contents of mental states. If I think that 'water is not fit to drink' it does not follow that I think that 'H_2O is not fit to drink', for I may not know that water is H_2O. (A transparent or de re construal would class these two beliefs as equivalent because they refer to 'the same thing').

Technically, the mark of an opaque (oblique) construal of a propositional attitude is either that the noun in the content clause is not substitutable with co-extensive expressions salva veritate, (hence when content clauses are not so substitutable they indicate different contents), or existential generalisation for this term is not a straight-forwardly valid transformation (Santa Claus).

Burge's point centres upon a thought experiment in which we, at first, consider a person who has a large number of beliefs and attitudes about 'arthritis', where the latter occurs obliquely in content clauses:

> For example, he thinks (correctly) that he has had arthritis for years, that his arthritis in his wrists and fingers is more painful than his arthritis in his ankles, that it is better to have arthritis than to have cancer of the liver, that stiffening joints is a symptom of arthritis ... (and) In addition to these unsurprising attitudes, he thinks falsely that he has developed arthritis in his thigh.[21]

But suppose (counterfactually) that this person had exactly the same life history, thoughts, internal monologues, sensations, behaviour, autonomously described mental states and so on (say he has a doppelganger on Twin Earth, or is himself quick frozen and taken to Twin Earth), yet the meaning of 'arthritis' was broadened to be inclusive of various other rheumatoid ailments, including those to be found in the thigh. That is, the only difference between the actual and counterfactual case is the subjects social environment; the accepted use and definition of the word 'arthritis". Burge then contends that:

> we cannot correctly ascribe any content clause containing an oblique occurrence of the term 'arthritis'.

---

[21] *op. cit.,,* p.77.

because;

> 'arthritis' in the counterfactual situation, differs both in dictionary definition and in extension from 'arthritis' as we use it.[22]

In other words, as in the Twin Earth examples, two different occurrences of the one term denote phenomena with two different extensions, thus they connote different contents. Because the extensions differ, content clauses containing 'arthritis' in the counterfactual situation are not freely substitutable salva veritate with arthritis - they are thus at the same time opaque, intensional, and constitutive of an individual's subjective representation - as opposed to an objective relation to an actual item or phenomenon in the world. The representationalist's tactic of using opaque construals of propositional attitudes in order to capture differences in the way a set circumstance may be 'taken' or 'interpreted' thus ensures that the mere correct attribution of a content sentence (or corresponding 'symbolic expression(s)) of the form (or 'shape') 'I have arthritis' does not thereby confer uninamity of content - and this on the relatively down to earth and non science-fictional understanding that different sectors of the community may take certain terms to have different extensions. If a determinate set of symbolic codes do not by the integrity of their identity (their intension) determine a discrete *semantics,* then we cannot suppose that the *meanings* of inputs are except contingently an indicator of 'symbolic' change, and thus of non-architectural change, and thus of the absence of architectural change, and thus of the absence of analogue process.

---

[22] *ibid.*, p.79

V. SUMMARY AND CONCLUSION.

The above results impugn the qualities of a determinate and intrinsic content upon which Pylyshn's criticism of the analog theory relies. As the semantics of a mental representation have to do with what it is true of, so the alterations of reference whilst holding autonomous psychological state state constant ensures that the latter changes in its content. If it were a belief, for instance, in the counterfactual situation it would become a different belief. In the case of sententially attributed beliefs with no clear or determinate content it is seen to be difficult to credit particular content sentence tokens with the full meaning their English language counterparts possess. In this we see a holism of belief content which contradicts the discrete sentence (or 'expression')content paradigm implied both by Pylyshn's theory and the Folk Paradigm to which it is wed.

The opacity of belief sentences is seen to connote representationality and the differential subjective 'takings' of sentences and situations necessary for the explanation of behaviour in these terms, whilst at the same time ensuring that a change in reference of a representation is thereby a change in extension and so a change in content. In short, representationality engenders indeterminacy of content and thus indeterminacy of representation. This, I think, is the fundamental paradox of representational psychology. To individuate belief states transparently, of course, is to say nothing, for my belief that my next-door neighbour is a nice man could become synonymous with the belief that a murderer and rapist is a nice man, which would fail to explain my friendly behaviour towards him. But to instantiate that belief in a tangible physical structure such as a mental sentence inscription or computational state is simultaneously to allow for variations in the content of that state which are ipso facto unexplained by that state. But the moral here is clearly to view representations as things which can possess contingent properties - so that they can be objects of scientific study - and a symbolic or intentional quality which varies over and above changes in the physical state of the organism is not a contingent property, being determined necessarily by

whatever the actual reference of that state might be. It is for instance a contingent question as to whether physical structure x can adequately represent for the organism domain y (e.g. the geometrical inter-relationships of physical objects), but not that structure x has semantics y, or reference y, for the semantics of a representation just are what it refers to.

Extrapolating from the thought experiments we can see that in the paradigmatic situation where one autonomous psychological state is attributed two or more different contents, by the non-equivalence of each mental-state extension pair with the other pair, the content of neither state is equivalent with the autonomous state. Autonomous states then, whilst they may be attributed contents 'from the outside', do not themselves have contents. If such states are alogarithmic and hence formal and syntactic we can follow Stich (1983) in terming the picture that emerges for representation the Syntactic Theory of Mind:

> The STM views mental states as relations to tokens of purely syntactic objects. Generalisations detailing the interactions of mental state tokens describe them in terms of their syntactic types. On the matter of content or semantic properties the STM is officially agnostic.[1]

Content is thus not excluded, but it is precluded as an explanatory factor of itself. Kosslyn's model is a paridigmatic syntactic theory. This is clear both from the explicit computational form in which the theory is expressed, and the almost implicit and by-the-way manner in which he seeks not to use the content of representations in explanation, but to reconstruct such content as the processes postulated may be ascribed, in terms of their quantifiable information processing qualities. The Turing machine example is a parallel: the actual function attributed to the machine qua mechanism is not present 'in' the machine in the form attributed. Quite how I don't know, but as long as the psychological moral remains of studying tangible alogarithms at the expense of blanket content ascriptions like 'tacit knowledge' then this is a *mere* philosophical question. The function of Philosophy in Cognitive Psychology we might say, is to do it - then set it aside, for only once it has been done does one become aware of how infected with philosophical prejudices one's

[1] Stich (1983), p.185.

previous 'a-philosophical' position was. Having put aside the deflationary theme of cognitive penetrability, the empirical imagery literature may be judged in terms of its ability to predict and integrate a domain unhindered by a priori counter-arguments.

If we reason this way however,the question emerges as to the interpretation of Pylyshn's apparently sound arguments for representation and the utility of mental content above. It is fairly easy to understand that aspect of the STM which renders the above examples innocuous: If mental states are individuated formally by physically instantiated computational/alogarithmic states then identity of narrow or autonomous state confers identity of psychological state. The unruly construct of content ceases to be cited *as* an explanation, thus ensuring that the philosophical minutae of content variations due to strange circumstances are no longer of significance to the psychological theory itself, except insofar as they militate against a particular construal of the computational model. But the question remains, do we not lose essential generalizations by adopting such a restrictive ontology of mental states ?

> It simply will not do as an explanation of, say, why Mary came running out of the smoke filled building, to say that there was a certain sequence of expressions computed in her mind according to certain expression transforming rules. ... The only way to both capture the important underlying generalizations ... *and* to see her behaviour as being rationally related to certain conditions is to take the ... step of interpreting the expressions in the theory as goals and beliefs.[2]

For familiar reasons, would not the formal/syntactic descriptions of a number of people in the 'help, the building is on fire' situation have nothing in common and so fail to explain their behaviours as instances of the belief state in question ? But Pylyshn, I think, begs the question, for there may well *be* nothing properly psychological in common amongst all occasions of fire avoidance behaviour. Our clue here is Theoretical Explanation as characterized above. Upon the construal of explanation I wish to emphasise, theories explain epistemologically immediate phenomena by, roughly speaking, characterising what the latter *really are*

---

2 Pylyshn (1980b), quoted in Stich (1983), p.171.

hence the fallibility of the immediate folk conception of psychological states is implicit in the attempt to explain them. It is not incumbent upon scientific theories to duplicate the categories of common sense. In fact, according to Science, common-sense is radically mistaken about a number of things; the solidity of every-day objects, their possession of colour, the 'gravitational attraction' between the Earth and a pencil tossed from the window.etc. (cf Churchland (1979)). To take an example, the folk-psychological state of 'remembering my fifth birthday'; I may remember my fifth birthday in a number of psychologically distinct ways. I may bring to mind a single image of a particular scene which occurred that day. I may recall a number of distinct but fragmentary images. I may recall certain sounds or smells. I may assume a state in which I can answer questions about what happened that day with out any accompanying fragmentary. Or I may simply make an appropriate one or two word verbal response to a question like 'did it rain on your fifth birthday ?' This is of course a very Wittgensteinian point. What reason do we have to take it for granted that all instances of what the ordinary man-in-the-street may choose to identify as 'reading' or 'understanding' *must* have a cognitive element in common? This is an empirical question which cognitive psychology may eventually settle. But the direction of informed consequence should be in favour of Science rather that Common sense. Thus whilst it is likely that most instances of running in fear from a burning building will turn out to share certain belief-like formal structures which could be interpreted along ordinary lines as something akin to sentences of mentalese, if psychology fails to find any such convergence then so much the worse for 'common sense'.

Pylyshn continues:

> What in the (syntactic) theory corresponds to this common interpretation ? Surely one cannot answer by pointing to some formal symbols. The right answer has to be something like the claim that the symbols represent the belief that the building is on fire -i.e., it is a semantic interpretation of the symbols as representing something. Otherwise even the relevance of the other symbols stored in memory would have to be gratuitously stipulated. If they were interpreted, on the other hand, we would see that they are relevant because they constitute premises which sanction the inference from

> such conditions as the smell of smoke to the belief directly responsible for the action.[3]

But this is a curious remark, for at the 'folk' level the content - which is incommensurable with physical or formal structure, as we have seen - of whatever belief and desires we may ascribe, manifestly relevant. But we do not thereby have to ascribe a specific content to each syntactic/representational 'symbol' any more than we must ascribe a distinct colour to each of the atoms which make up this tomato, or a distinct solidity to each of the atoms which compose this table, or a distinct gravitational attraction between each part of this ball and the Earth (cf General Relativity), despite the fact that the constituents of these things explain their overt qualities.

Beliefs, desires and Folk Psychology are a product, I think, of what Dennett calls 'the intentional *stance*':

> a particular thing is an intentional system only in relation to the strategies of someone who is trying to explain and predict its behaviour.[4]

> One predicts behaviour in such a case by ascribing to the system the *possession of certain information* and supposing it to be *directed by certain goals* , and then by working out the most reasonable or appropriate action on the basis of these ascriptions and suppositions.[5]

However;

> the definition of intentional systems ... does not say that intentional systems *really* have beliefs and desires, but that one can explain and predict their behaviour by *ascribing* beliefs and desires to them.[6]

---

[3] Pylyshn, quoted in Stich, *op. cit.*
[4] Dennett (1978) p.3-4
[5] *ibid*, p.6
[6] ibid, p.7

The question, then, of 'where and what is mental content ?' is, I think, the wrong question. It is like asking 'where is the solidity of this desk ?' Treating it as solid certainly explains its effect upon my foot if dropped on it, and at the everyday level this is enough. Treating cognition as a formal computational process has specific scientific virtues. It manifests, roughly speaking, the best approximation we have at the moment of what cognition 'really is'.

ANDERSON, J: 'Arguments Concerning Representations for Mental Imagery'. *Psychological Review* , 85: 249-77, 1978.

ARBIB, M.A: *Theories of Abstract Automata* . Englewood Cliffs, N.J: Prentice Hall, 1969.

ARISTOTLE *Metaphysics* . Warrington, J., (ed. & trans.) London: Dent, 1966.

ARMSTRONG, D.M: *The Nature of Mind, and other essays*. St. Lucia, Queensland: University of Queensland Press, 1980.

BRAND, G: *The Central Texts of Wittgenstein* . Oxford: Blackwell, 1979.

BRENTANO,F: *Psychology From An Empirical Standpoint* . (1874). McAlister, L.C., Rancurello, A.C., Terrell, D.B., (trans.). London: Routledge and Kegan Paul, 1973.

BURGE,T: 'Individualism and the Mental'. in French,P., Uehling,T., Wettstein,H.,(eds) *Midwest Studies in Philosophy, vol. 4, Studies in Epistemology*. Minneapolis: University of Minnesota Press, 1979.

BURGE, T: 'Individualism and Psychology'. *Psychological Review*, 95: Reprinted in Silver,S, (ed.)

BUXTON, C. E: (ed) *Points of view in the modern History of Psychology*. New York: Academic Press.

CHURCHLAND, P.M: *Scientific Realism and the Plasticity of Mind* . Cambridge: Cambridge University Press, 1979.

COOPER, L.A: 'Mental rotation of random two dimensional shapes'. *Cognitive Psychology*, 7: 20-43, 1975.

DAVIDSON, D: 'Mental Events' in Honderich, T., Burnyeat, M., (eds.) *Philosophy As It Is* . Middlesex: Pelican, 1979. p. 218-38.

DENNETT,D: *Brainstorms*, Cambridge,Mass: MIT Press.

DESCARTES, R: *Selected Philosophical Writings* . Cottingham, J., Stoothoff, R., Murdoch, D., (trans.). Cambridge: Cambridge University Press, 1988.

DEWEY, J: *Experience and Nature.* Chicago: Open Court, 1925.

EDWARDS, P: (editor in chief) *The Encyclopedia of Philosophy* . New York: Macmillan, 1967.

FEYERABEND, P: *Problems with Empiricism.* Cambridge: Cambridge University Press, 1981.

FINKE, R,A., SCHMIDT, M.J: 'Orientation-specific colour after effects following imagination'. *Journal of Experimental Psychology: Human Perception and Performance*, 3: 599-606,1977.

FODOR, J.A: *Psychological Explanation* . New York: Random House, 1968.

FODOR, J.A: 'The Appeal to Tacit Knowledge in Psychological Explanation'. *Journal of Philosophy* , 65: 627-40, 1968.

FODOR, J.A: *The Language of Thought* . New York: Crowell, 1975.

FODOR, J.A: 'Tom Swift and His Procedural Grandmother'. *Cognition* , 6: 229-47. 1978.

FODOR, J.A: 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology'. *Behavioural and Brain Sciences* , 3: 63-109, 1980.

FODOR, J.A., PYLYSHN, Z.P: 'How Direct is Visual Perception ? Some Reflections on Gibson's "Ecological Approach"'. *Cognition* , 9: 139-96, 1981.

FRAISSE, P: *The Psychology of Time* . New York: Harper and Row, 1963.

GOODMAN, N: *The Language of Art: An Approach to a Theory of Symbols.* London: Oxford University Press, 1969.

HEMPEL, C.G: *Philosophy of Natural Science.* Englewood cliffs, N.J: Prentice Hall, 1966.

HOFSTADTER, D.R: *Godel, Escher, Bach: An Eternal Golden Braid* . Hassocks: Harvester Press, 1979.

KIM, J: 'Supervenience and Nomological Incommensurables'. *American Philosophical Quarterly* . 15: 149-56,1978.

KOSSLYN, S.M: 'Information Representation in Mental Images'. Cognitive Psychology, 7: 341-70, 1975.

KOSSLYN, S.M: 'Can Images be Distinguished From Other Forms Of Internal Representation: Evidence From Studies of Information Retrieval Time'. *Memory and Cognition* , 4: 291-97, 1976.

KOSSLYN, S.M., *Image and Mind.* Cambridge, Mass: Harvard University Press, 1980.

KOSSLYN, S.M: 'The Medium and the Message in Mental Imagery: A Theory'. *Psychological Review* , 88: 46-66. 1981.

KOSSLYN, S.M., BALL, T.M., REISSER, B.J: 'Visual Images Preserve Metric Spatial Information: Evidence From Studies of Image Scanning'. *Journal of Experimental Psychology: Human Perception and Performance* 4: 47-60, 1978.

KOSSLYN, S.M., PINKER, S., SMITH, G.E., SCHWARTZ, S.P: 'On The Demystification of Mental Imagery', *Behavioural and Brain Sciences* , 2: 535-81, 1979.

McDERMOTT, H.J.S: The Writings of William James. New York: Random House 1967.

NAGEL, E: *The Structure of Science.* London: Routledge and Kegan Paul, 1961.

NEWELL, A., SIMON, H.A: *Human Problem Solving* . Englewood cliffs, N.J: Prentice Hall, 1972.

PEARS, D.F. (ed), *The Nature of Metaphysics,* London: Macmillan, 1957. 1962

POPPER, K: *The Logic of Scientific Discovery.* 2nd. ed. London: Hutchison, 1959.

POSNER, M: *Chronometric Explorations of Mind* . Hillsdale, N.J: Erlbaum, 1978.

PUTNAM, H: 'The Meaning of "Meaning"' in *Philosophical Papers Vol. 1:Mind, Language and Reality.,* Cambridge: Cambridge University Press.

PYLYSHN, Z.P: 'Computational Models and Empirical Constraints'. *Behavioural and Brain Sciences* , 1: 93-127, 1978.

PYLYSHN, Z.P: 'Computation and Cognition', *Behavioural and Brain Sciences,* 3: 111-69. 1980.

PYLYSHN,Z.P: 'Cognitive Representation and the Process Architecture distinction'. *Behavioural and Brain Sciences,* 3: 154-69, 1980.

PYLYSHN, Z.P: 'The Imagery debate: Analogue Media versus Tacit Knowledge'. *Psychological Review* , 88: 16-45, 1981.

PYLYSHN, Z.P: *Computation and Cognition.* Cambidge, Mass: MIT Press, 1984.

REED, E.S: 'Descartes Corporeal Ideas Hypothesis and the Origin of Scientific Psychology'. *Review of Metaphysics* , 35: 731-52, 1982.

RICHARDSON, J.T.E: *Mental Imagery and Human Memory* . New York: St. Martins, 1980.

ROSS, L: 'The Intuitive Psychologist and his Shortcomings' in Berkowitz, L., (ed.) *Advances in Experimental Social Psychology . Vol. 10* . New York: Academic Press, 1977.

ROSS, L., LEPPER, M., HUBBARD, M., 'Perseverance in Self Perception and Social Perception: Biased Attributional Processes in the

Debriefing Paradigm'. *Journal of Personality and Social Psychology* . 32: 880-92, 1975.

RYLE, G: *The Concept of Mind.* London: Hutchison, 1949.

SELLARS, W: *Science, Perception and Reality.* London: Routledge and Kegan Paul, 1963.

SELLARS, W: *Philosophical Perspectives,* Springfield, Ill: Charles C. Thomas, 1967.

SHAFFER, J.A: *Philosophy of Mind.* Englewood cliffs, N.J: Prentice Hall, 1968.

SHEPARD, R: 'Ecological Constraints on Internal Representation: Resonant Kinematics of Perceiving, Imaging, Thinking and Dreaming'. *Psychological Review* 91: 417-47, 1984.

SHEPARD, R., COOPER. L.A: *Mental Images and Their Transformations.* Cambridge, Mass: MIT Press, 1982.

SHEPARD, R., METZLER, J: 'Mental Rotation of Three Dimensional Objects', *Science* 171: 701-3,1971.

SILVERS,S: *Representation: Readings in the Philosophy of Mental Representation,* Dordrecht, Boston:Kluwer Academic Publishers, 1989.

SKINNER, B.F: "Are Theories of Learning Necessary ?'. *Psychological Review,* 57: 193-214. 1950.

SKINNER, B.F: 'The Concept of the Reflex in the Description of Behaviour'. Journal of General Psychology,, 5: 427-58, 1931.

SKINNER, B.F: The Shaping of a Behaviourist, New York: Knopf, 1979

SKINNER, B.F: *Verbal Behaviour.* New York: Appleton-Century-Crofts, 1957.

SMART, J.C.C: *Philosophy and Scientific Realism.* London: Routledge and Kegan Paul, 1963.

STICH, S.P: 'Autonomous Psychology and the Belief-Desire Thesis'. *The Monist* , 61: 573-591, 1978.

STICH, S.P: *From Folk Psychology to Cognitive Science* . Cambridge, Mass: MIT Press, 1983.

TURING, A: 'On Computable Numbers, with an Application to the Entscheidungs Problem'. *Proceedings of the London Mathematical Society* , Series 2: 230-65, 1936.

WATSON, J: 'Psychology as the Behaviourist views it'. *Psychological Review,* ,20: 158-77. 1913.

WITTGENSTEIN, L: *Philosophical Investigations*. Oxford: Blackwell, 1953.